

Modeling Route Choice of Utilitarian Bikeshare Users from GPS Data

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Ranjit Khatri
December 2015

Copyright © 2015 by Ranjit Khatri
All rights reserved.

DEDICATION

To My Late Grandfather

ACKNOWLEDGEMENTS

It is a great pleasure to acknowledge my sincere gratitude to my advisor, Dr. Christopher R. Cherry, for his continuous support during my graduate studies and related research. His guidance, motivation, and patience have helped me in all the time of research and writing of this thesis.

I would like to thank Dr. Lee D. Han and Dr. Shashi S. Nambisan, who agreed to be a part of my thesis committee. Their insightful comments and encouragement contributed a lot in improving my thesis.

I would also like to thank Southeastern Transportation Center (STC), a Federal University Transportation Center, for funding this project and Social Bicycles for providing the GPS data from the Grid Bikeshare – Phoenix, AZ.

I am very grateful to my lab mates for the help during my study at The University of Tennessee, Knoxville. They were always willing to help during my study and offer suggestions regarding my future plans.

Finally, I would like to thank my family: my parents and my sister for supporting me throughout my life in every good thing I do.

ABSTRACT

This research examines the behavior of bikeshare users from Grid Bikeshare Program in Phoenix, Arizona under two behavioral frameworks: facility usage assessment and route choice assessment. The analysis is performed for the two different categories of subscribers: registered and casual subscribers. This is the first study that uses the real-time GPS data from bikeshare users to model their route preferences. The data used for this study were obtained from 9,101 trips made by 1,866 bikeshare. An important aspect of this bikeshare is that it allows non-station origin and destinations. The GPS points collected from the trips made by bikeshare users were matched to the street base network to determine the attributes of the route followed by the cyclists. Facility usage assessment included the determinations of use of roadway segments based on Annual Average Daily Traffic, posted speed limit, and roadway classification. Similarly, wrong direction riding behavior on the road was compared for one-way versus two-way roads and road segments with bicycle facilities versus without bicycle-facilities. Route choice decisions were modeled using the Path Size Logit model, which is based on a Multinomial Logit framework. The major findings include behavioral differences between the two groups of users such as average distance travelled, time of the day and day of the week variation and composition of the total users. Registered users, although fewer in number, made significant number of trips. Casual users were involved more in wrong direction riding in forty selected road segments from Downtown of Phoenix. The results from the discrete route choice model show that riders were very sensitive to travel distance, with positive utility towards using bike-friendly infrastructure. Having bike-specific infrastructures for the complete route is equivalent to decreasing distance by 44.9% (53.3% for casual users). Left turns imposed higher disutility for casual users as compared to right turns. A number of signalized intersections had a positive effect in selecting the route whereas the proportion of one-way segments, traffic volume and length of the route had a negative influence on route choice.

TABLE OF CONTENTS

1. INTRODUCTION.....	1
1.1. Research Background	1
1.2. Research Objectives	2
1.3. Study Area	3
2. LITERATURE REVIEW	4
2.1. Route Choice Criteria Based on Attributes of Traversed Road	4
2.2. Impact of Cyclists' Behavior and Characteristics on Safety	5
2.3. New Sources of Bicycle Data: Phone App users and Bikeshare users	7
2.4. Advantages and Challenges Associated with GPS Data.....	9
3. DATA DESCRIPTION	11
3.1. Data Sources	11
3.1.1. Bikeshare GPS Data	11
3.1.2. Road Network.....	12
3.2. Data Cleaning.....	12
3.3. Completing The Road Network	14
4. CONCEPTUAL FRAMEWORK AND METHODOLOGY	15
4.1. Map Matching.....	15
4.1.1. Available Methods.....	15
4.1.2. Used Procedure.....	16
4.1.3. Issues and Solutions.....	17
4.2. Choice Set Generation	20
4.2.1. Removing the Identical Alternatives	22
4.2.2. Calculating the Attributes of Alternatives	22

4.3.	Discrete Route Choice Model	23
4.3.1.	Assumptions of Multinomial Logit Model	24
4.3.2.	Path Size Logit Model	25
4.3.3.	Distance Trade-off Calculation.....	27
4.4.	Variables Utilized in the Model and Hypothesis of Utility.....	27
5.	ANALYSIS AND RESULTS	29
5.1.	Demographics of the Bikeshare Users	29
5.2.	The Wrong Direction Riding	31
5.3.	Facility Usage Assessment.....	34
5.3.1.	Distribution of Different AADT Levels or Speed Limits on the Observed Route ..	34
5.3.2.	Distribution of Different Roadway Infrastructures on the Observed Route	36
5.4.	Route Choice Assessment.....	37
6.	DISCUSSION	40
7.	CONCLUSIONS, LIMITATIONS, AND FUTURE RESEARCH	45
	LIST OF REFERENCES.....	47
	APPENDIX.....	51
	VITA.....	56

LIST OF TABLES

Table 1 Percentage of Trips with Different Number of Alternatives	22
Table 2 Variables Used for the Choice Model and Corresponding Hypothesis	28
Table 3 Estimation of Utility Coefficients.....	39
Table 4 Distance Value (%) for Unit Change in Attribute	39

LIST OF FIGURES

Figure 1 Possible Recreational Bikeshare trip	12
Figure 2 Result of the ArcGIS Model Used for Map Matching of GPS Points.....	17
Figure 3 Inability to Merge Buffers due to Sparse Consecutive Points.....	19
Figure 4 (a) Failure to Predict a Trip with Loops (b) Missing Links in the Trip.....	19
Figure 5 Distribution of Registered and Casual Subscribers	29
Figure 6 Time of the Day Variation of Percentage of Trips for the Bikeshare Users	30
Figure 7 Day of the Week Variation of Number of Trips for the Bikeshare Users	31
Figure 8 Wrong Direction Riding for One-way and Two-Way Road Segments.....	33
Figure 9 Wrong Direction Riding on the Segments With or Without Bike Facilities	33
Figure 10 Distribution of Travel Distance Over Range of AADT	35
Figure 11 Distribution of Travel Distance Over Range of Speed Limit.....	35
Figure 12 Usage of Different Bicycle Facilities	36
Figure 13 Usage of Different Roadway Infrastructures.....	37
Figure 14 Categories of Bicycle-Specific Facilities in Phoenix, AZ.....	52
Figure 15 Street Segments Used for Analyzing Wrong Direction Riding Behavior	53
Figure 16 Set of Alternatives for a Pair of Origin and Destination	54
Figure 17 Volume of Bikeshare Trips on Streets of Downtown Phoenix	55

1. INTRODUCTION

1.1. Research Background

The bicycle has been increasingly used as a means of transportation in the United States and other countries of the world, considering all the major advantages of this mode. Along with the growing concerns towards sustainable mode of transportation, use of bicycle as an alternative mode of transportation could be a better solution to some of the problems like dependency on automobile, lack of parking, increased greenhouse gases and so on.

For successful modal shift to bicycle, two things should be done. First, inducing new users to cycle instead of auto travel. Second, motivating the continued use of cycling by assuring them the reliability and safety of this mode of transportation. Proper bicycle route planning is the foremost step to achieve these objectives, which should be driven by the detailed analysis of cyclists' behavior. The way in which cyclists interact with the roadway and traffic characteristics is directly associated with safety. Analyzing route choice behavior is essential to keep them safe on the road. Furthermore, facility usage assessment of the cyclists could assist in assigning the right facilities in the right location.

With the booming use of smartphones, people are using smartphone applications such as route tracking applications and/or fitness tracking applications to record and track their data. Similar applications are being used by the cities across the US, like Cycle Tracks (San Francisco, Calif.), Cycle Atlanta (Atlanta, Georgia), CyclePhilly (Philadelphia, Penn.) and I Bike KIX (Knoxville, Tenn.). By providing real-time GPS data to planners, these applications allow for disaggregate analysis and inform transportation planning in these cities.

In the last decade, bicycle-sharing systems have gained popularity in many North American cities along with the other major cities in the world. Most of the trips are made for the short distance point-to-point travel thus improving accessibility in cities. With a

large number of bikeshare systems being in effect, analysis of their behavior is an important task. With GPS devices embedded in the bicycles, bikeshare system can provide immense wealth of data for analysis of data. This dataset would provide the exact behavior of the bikeshare users to assess their decisions and behavior on the road in response to the information perceived by them.

1.2. Research Objectives

The research objectives of this study can be grouped into two general categories of behavior assessment: facility usage assessment and route choice assessment. Following are the research questions for both categories of behavior assessment.

1. Facility usage assessment

- i. How often do the bikeshare users ride against the traffic either in one-way or two-way road segments?
- ii. What is the distribution bikeshare trips over functional classification of road and different bicycle-specific infrastructures?
- iii. What is the distribution of bikeshare trips over different speed limits and AADT?

2. Route choice assessment

What is the extent of following factors that influence the decision to choose a certain route among the available alternatives?

- AADT along the route
- Bicycle-specific facilities
- Posted speed limit of the vehicle
- A number of signalized intersections.
- A number of left and right turns.
- Proportion of one-way streets on the route
- Time of the day and day of the week

The study and analysis are limited to the utilitarian bikeshare trips so as to remove any anomalous behavior introduced by a recreational trip that might skew the interpretation of the results. Additionally, the analysis are compared over registered subscribers and casual subscribers of the bikeshare. Comparing the behavior of these two groups is very important as they might behave differently on the road. Also, the results of this study are compared with previous studies on route choice behavior of conventional cyclists.

1.3. Study Area

The data was obtained directly from the bikeshare system “Grid Bikeshare” in Phoenix, AZ. The Grid Bikeshare system, which was launched in Fall 2014 with 27 stations, currently includes approximately 500 bikes and 39 stations (or hubs). The stations are placed on an area that is approximately 2.5 km East to West and 8 km North to South, covering downtown of Phoenix. Although the system relies heavily on stations, users can also park bikes away from stations for a small fee. This gives flexibility to the users regarding parking as well and possibly attract a high number of users. The target population for the study was all cyclists who either register monthly/annually for the bikeshare or are casual users who pay a marginal fee for renting a bike. This is a unique bikeshare system, which records GPS data for all of the trips, operates on a grid street network and allows non-station origins and destination. All of this enables unique route choice analysis of the bikeshare users. Bike lanes cover a large proportion of bicycle facilities in City of Phoenix, AZ [Figure 14 in Appendix]. In our study area, more than 75% of the paths are bicycle lanes.

2. LITERATURE REVIEW

2.1. Route Choice Criteria Based on Attributes of Traversed Road

Most of the transportation models in North America do not include bicycling in all the steps of transportation planning, and, if included, it is generally assumed that cyclists choose the minimum path between origin and destination and travel with fixed travel speed (Larsen & El-Geneidy, 2011). In most of the cases, route choice of the automobiles solely depends on the shorter length of the route and lesser duration of travel. Contrary to this, this might not be always true for the cyclists; selection of route depends on the distance, safety, turn frequency, slope, intersection control, and traffic volumes (Broach, Dill, & Gliebe, 2012; Broach, Gliebe, & Dill, 2009; Ehrgott, Wang, Raith, & Van Houtte, 2012; Hood, Sall, & Charlton, 2011; Sener, Eluru, & Bhat, 2009; Winters, Davidson, Kao, & Teschke, 2011). In short, cyclists have two objectives while selecting the route, i.e. travel time and suitability of the route (Ehrgott et al., 2012).

There are two main challenges associated with increasing cycling: inducing new cyclists and motivating current cyclists to continue cycling. There are many studies that suggest that bicycling facilities induce new cyclists, in addition to encouraging the existing cyclists to cycle (Menghini, Carrasco, Schüssler, & Axhausen, 2010; Sener et al., 2009). The relative effect of the various facilities, however, is contradictory for many studies. Some study found bike lane superior to other bike facilities (Hood et al., 2011). Another study by Broach et al. (2012) found off-street bike paths were valued more than other facilities (Broach et al., 2012) while another found one-way bicycle path was safer than bicycle lane (Schepers, Kroeze, Sweers, & Wüst, 2011). In another study, wide outside lane and bike lanes were compared and it found bike lanes to be safer (Duthie, Brady, Mills, & Machemehl, 2010). Accurately determining the relative effect of these facilities require further detailed examination considering the possible effects of other relevant factors like traffic volume, the length of the facility, and so on. Also, bicycling along the one-way street acts as disutility to cyclists because cyclists do this only when

the activity saves more than four times the distance (Hood et al., 2011). This demands the construction of bicycle lanes in those areas which are compelling riders to travel more due to perceived risk associated with this mode.

Different from other studies, the study (Larsen & El-Geneidy, 2011) focuses on ascertaining the influence of spatial characteristics (length, width) of specific bicycle facilities. This study found that the longer the facility, the greater the possibility that they will deviate longer distance to use the facilities. Proximity to bicycle facilities was another factor to promote and increase bicycling (Larsen & El-Geneidy, 2011).

In addition to all these, land use is another major factor that has a direct influence on how often people cycle for the utilitarian trips. People are motivated to bicycle when the time required to get to the multiple destinations is comparatively less such that it outweighs the advantages of the automobile. For the areas having a mixed land use, with the higher connectivity of networks, shorter trips are possible. This can promote cycling among the people (Dill & Gliebe, 2008).

2.2. Impact of Cyclists' Behavior and Characteristics on Safety

While roadway factors constitute many factors that can influence the route preferences of the cyclists, characteristics and behavior of cyclists also have significant impact in route choice and safety.

Attitude and perception of road users—both cyclists and drivers—is another major factor that defines the safety on the road. Cyclists, who enter the intersection, expect that they would be given right of way. Most of the drivers of the vehicle looked for other vehicles that could conflict with their path but failed to see the cyclists, and this often results in a bicycle crash (Räsänen & Summala, 1998). In the same study, it was found that 68% of cyclists noticed that drivers were approaching and 92% of them, who noticed the drivers, thought that the drivers would give way as required by law (Räsänen & Summala, 1998). However, this is not the main problem for other countries, like the

Netherlands, where people have adapted to scan for cyclists in the road due to a large number of cyclists (Schepers et al., 2011). Besides, most of the drivers as well as cyclists believe that using some aids (fluorescent vests, reflectors, etc.) improved their visibility compared to normal clothing (Wood, Lacherez, Marszalek, & King, 2009).

Another important facet of route choice is rider characteristics, like the level of experience and their extent of cycling. Some claim that inexperienced cyclists are more inclined to use bicycle facilities compared to experienced ones (Larsen & El-Geneidy, 2011), whereas some found experienced cyclists preferred the bicycle facilities (Broach et al., 2012). Similarly, infrequent (inexperienced) cyclists, tend to have strong preferences for bicycle lanes (Hood et al., 2011). The faster cyclist, who are generally experienced, was found to prefer marked routes (Menghini et al., 2010).

Route selection depends on the trip purpose if it is a commute trip or non-commute trip. For the commute trips, cyclists are more attracted to shorter routes, they are not sensitive to bicycle infrastructure on the route (Broach et al., 2012; Broach et al., 2009; Dill & Gliebe, 2008), and avoid hills (Broach et al., 2009; Dill & Gliebe, 2008; Hood et al., 2011). The possible explanation for this might be that commuters are more aware of the route and know that route very well. Hence, they make a travel such that it reduces travel time and effort.

As presented by previous literatures, a large number of crashes in the intersections poses a threat to the safety of the cyclists. Numerous studies (e.g. (Broach et al., 2009; Menghini et al., 2010)) have found that cyclists avoid signalized intersections and stops. This behavior is more pronounced among commuters as compared to non-commuters (Broach et al., 2009). Although the rate of infringement of the traffic signal, which was found to be seven percent, increases when the traffic volume is reduced to a level perceived by the cyclists as a safe option, it is still a dangerous behavior (Johnson, Newstead, Charlton, & Oxley, 2011). However, this rate of infringement was more for the cyclists turning right as compared to cyclists going straight. Furthermore, the rate of

infringement was higher for un-signalized intersections than signalized intersections and it was dependent on the slope and intersecting traffic volumes (Langford, Chen, & Cherry, 2015).

2.3. New Sources of Bicycle Data: Phone App users and Bikeshare users

Planners require good bicycling data – preferably not suffering from self-selection bias – which can be used to understand the behavior of cyclists. This requirement is complemented by new methods of real time data collection using dedicated GPS devices or built-in GPS in smartphones. This has facilitated researchers and practitioners with new techniques to assess the route choice and behavior of cyclists on the road. However, these types of data suffer from self-selection bias, i.e., users opt to use an app to record their trip and know much about riding efficiently in most of the case. Some cyclists use smartphone applications such as Strava, MapMyRide, CycleMaps or other fitness tracking applications to record and track their data in order to encourage physical activity and healthy living (Klasnja & Pratt, 2012). However, most of those data sources are not usually accessible to planners.

Leveraging this technology, some cities are utilizing GPS data collection techniques from open source applications like Cycle Tracks (Hood et al., 2011). These data collection techniques utilize built-in GPS capabilities of smartphones, which provides high quality revealed data at a reduced cost compared to stated preference surveys. The collected data is directly sent to remote servers without any requirement to go to the field to retrieve the data. There are several applications that are being used in cities of US, like Cycle Tracks (San Francisco, Calif.), Cycle Atlanta (Atlanta, Georgia), CyclePhilly (Philadelphia, Penn.), My ResoVelo (Montreal, Quebec), and I Bike KNX (Knoxville, Tenn.). The data from these apps can inform transportation planning in these cities and allows for disaggregate analysis. Nevertheless, one of the challenges with app-based data collection is that users have to opt-in and use the application for every trip.

In the last decade, the bikesharing system has gained popularity in many North American cities along with the other major cities in the world. There are more than one million bicycles under the bikesharing scheme in more than 500 cities of 49 countries. These bikesharing schemes allow individuals to use a bicycle for a certain period between fixed bikeshare stations (The Bike-Sharing World Map). Some, like Grid Bikeshare in Phoenix, Ariz., have facilitated the use of public racks as the bike stations too. Availability of bikeshare is meant for efficient short distance travel, thus solving the “first/last mile problem” by connecting to other modes or providing urban circulation.

Although bikeshare is meant to be beneficial in reducing car use and increasing bicycle trips, some results suggest that bikeshare replaces most public transit trips and walk trips rather than car trips (Shaheen, Guzman, & Zhang, 2010). In addition to expanding docking stations and making convenient use of bikeshare, high substitution of car trips could be only obtained by making the travel time of bikeshare trips competitive to that of car trip by achieving efficient routing or improving bicycle amenities (Fishman, Washington, & Haworth, 2014).

Bikeshare systems are ripe for developing new data streams to understand bicycling behavior in cities. Several recent studies have mined bikeshare data to understand flows between stations and identify differences in user types. Bikeshare users are generally classified as a registered users (frequent users who subscribe to a membership that usually includes unlimited use for the duration of the membership) and casual users (occasional users who pay for service as they use it, often travelers or tourists). Unlike the casual subscribers, who primarily make recreational trips, commuting is the main purpose for registered subscribers (Fishman, Washington, & Haworth, 2013).

Most of the previous literature on bikeshare users focus on the demographics of users (Buck et al., 2013), or station or system performance (Schoner & Levinson, 2013). Recent bikeshare systems have included vehicle tracking telematics onboard the bicycle,

which allows for a finer level of analysis, i.e., vehicle level of analysis instead of station level of analysis. This has opened a new opportunity to investigate route choice, particularly as it relates to safety and comfort, of an entire sub-population of cyclists, bikeshare users. This subpopulation is an important group because it constitutes a large portion of urban cyclists and represents an important part of the travel trip, generally short urban center trips.

2.4. Advantages and Challenges Associated with GPS Data

In the course of understanding the riding behavior of cyclists, several efforts have been made to determine route choice behavior. Two main approaches to explain route choice behavior of cyclists hinge on either stated preference (SP) data (Segadilha & da Penha Sanches, 2014; Sener et al., 2009; Stinson & Bhat, 2003) or revealed preference (RP) data (Broach et al., 2009; Dill & Gliebe, 2008). Most of these studies have focused on the presence of various bike-specific infrastructure, route attributes, individual characteristics, land use and so on. There are numerous studies that use SP surveys because of the ease in collecting data, which is free from extraneous observations, and simplicity in modeling. Typical SP surveys allow the participants to rate different type of facilities and choose among different available options, or some SP surveys also might allow the people to recall the path they have followed from their memories and complete the information required by the surveys.

Most of these studies attempt to model behavioral intent and are inflicted by the possibility that responses might be biased from their actual behavior (Winters et al., 2011). Hence, the shortcomings possessed by the studies based on the data collected from these SP surveys could be eliminated by the use of GPS to collect the accurate route data as a Revealed Preference (RP) surveys. Another advantage of using GPS for RP surveys is that, in addition to the reduction in burden on the participants to remember the route, it can accurately provide origin, destination, travel time, and travel route. Revealed Preference surveys have been used in many surveys: some of them recruited the

participants and obtained data from GPS logging devices, whereas some of them collected data from the built-in GPS of the mobile phone through a smartphone application.

Although with all these benefits over stated preference surveys, there are problems associated with this method of data collection. Taking the GPS activated devices, for instance, could deviate the behavior of the cyclists making rational decisions, which they would not have made in the absence of the GPS units with them. Having to activate GPS units or smartphone applications at the start of every trip might be considered too tedious, especially for short trips. Rapid battery depletion of mobile phone and requirement to charge GPS units are other barriers for GPS data collection.

Additionally, the quality of the GPS points collected differs from one GPS unit to other based on the quality of the device. The inaccuracy is more pronounced in terms of signal noise and signal loss. Furthermore, GPS data quality is highly affected by the location of GPS in built environment because it should have contact with at least four satellites to accurately locate the point on the earth. Tall buildings/urban canyon, for instance, could restrict the accurate data collection by obscuring the GPS signal reception. As a result, this may give the wrong information about the exact route taken by the cyclists. Despite the cheap and efficient data collection, in addition to aforementioned drawbacks, complicated data processing of huge bulk of data and accurately representing them on the detailed street network can be challenging.

3. DATA DESCRIPTION

The GPS data obtained from Social Bicycles and Street network obtained from Maricopa Associations of Governments are the two major data sources utilized in this study. These data acquired from respective agencies are cleaned and prepared for the further analysis.

3.1. Data Sources

3.1.1. *Bikeshare GPS Data*

The data was collected and stored by Social Bicycles in their repository, which was provided to us for the study. Bikeshare, with instrumented bikes, allows for better assessment of revealed route preference of a large sub-population of cyclists. According to the type of subscription, the users of the Grid Bikeshare are broadly categorized into two groups as registered subscribers (frequent cyclists who subscribe to a membership that usually includes unlimited use for the duration of the membership) and casual subscribers (occasional users who pay for service as they use it, often tourists). Hence, the dataset is segmented into two broad categories: Registered and Casual Subscribers.

The data consists of coordinates, at a sub-minute resolution for every trip that was made by the users of Grid Bikeshare users. The raw data provided to us by Social Bicycles also includes the date of travel, start and end time of the trip, total travel time and distance travelled. However, this data did not have timestamps and were not uniformly spaced in terms of time or distance. This restricted the study of the speed of travel at various roadway infrastructures. The frequency of the GPS readings varied from 1 per minute to 25 per minute. For this reason, the GPS readings were not spaced equally in terms of distance and time.

3.1.2. Road Network

The road network in a GIS environment was provided by the Maricopa Association of Governments. It included attributes for roadway segments that are of interest to this study e.g., Average Annual Daily Traffic (AADT), the geometry of the road, and bike-specific facilities. This network dataset was supplemented by a number of bicycle motor vehicle crashes and number of signalized intersections of the study area.

3.2. Data Cleaning

The first phase of data cleaning is done with an objective of removing two types of trips: very short trips and recreational trips. Figure 1 shows a typical recreational trip. Short trips do not have sufficient information for its assignment to street network. In most of the cases, shorter trips do not represent the actual trips as these kind of trips might have been recorded due to technical errors. This is supported by the presence of large number of trips with zero distance and travel time. This cleaning of trips will yield those trips for which trip assignment and behavior analysis is possible.

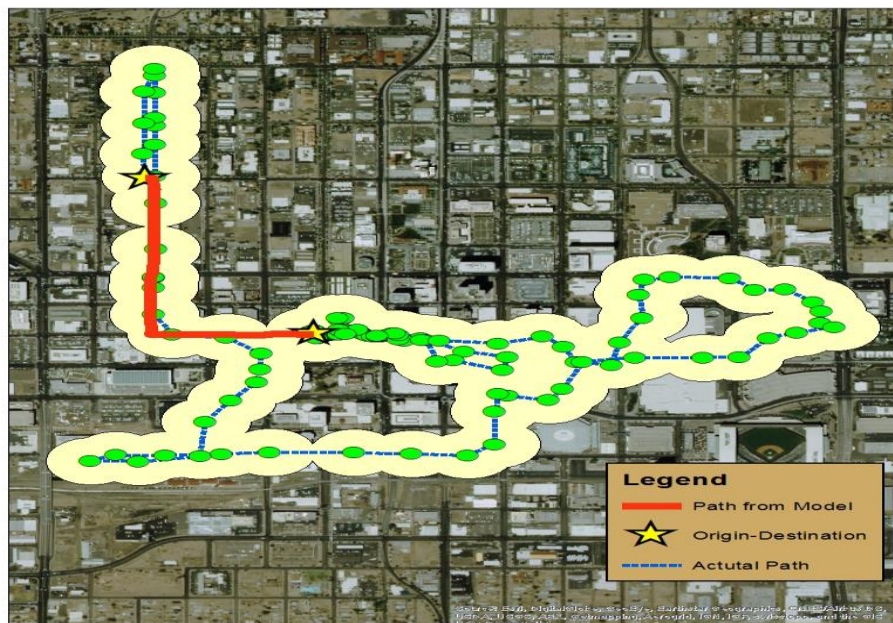


Figure 1 Possible Recreational Bikeshare trip

Another main objective of the data cleaning is to remove all the possible recreational trips. With a high number of trips made on weekends, it becomes necessary to remove possible recreational trips. This was done for the current scope of analysis of utilitarian bikeshare trips because bicycle trips for recreational purposes are very different from the utilitarian trips. For instance, recreational cyclists – without apparent destinations – might use longer route including bicycle specific facilities. Also, many recreational trips returned to the origin, or included loops, making route assignment and identification of alternate routes challenging.

The following are the basic criteria for the data cleaning process.

1. Trips that satisfied following threshold were removed.
 - a. Travel Time < 1 min
 - b. Travel Time > 90 minutes
 - c. Travel distance < 0.02 miles
 - d. Travel distance > 10 miles
 - e. Average velocity < 1.5 mph
 - f. Average velocity > 25 mph
 - g. Trips having fewer than 10 GPS points
2. Trips based on the threshold assigned with respect to origin-destination distance and shortest distance were removed to eliminate circuitous tours that were not likely destined for a specific place.
 - a. Trip distance > 3× O-D “as the crow flies” distance
 - b. Trip distance > 2.5× shortest possible travel distance between the O-D pair

These numbers were assigned based on the assumption that any person will not make the trip such that it is more than three times the O-D–”as the crow flies”–distance or more than 2.5 times the shortest possible travel distance between O-D pair.

3. Certain GPS points very close others were removed to reduce the volume of data required for analysis. A point that was at a distance of fewer than 10 feet from both the preceding and succeeding points was deleted.

There were 20,468 trips in the raw data. Using first criteria mentioned above, 3,925 trips (20% of 20,468) were removed. For the remaining 16,543 trips, criteria 2(a) removed approximately 25% of the remaining trips. There were only additional 71 trips deleted from the criteria 2(b), as most of the trips satisfying criteria 2(b) also satisfied criteria 2(a) and were previously removed.

There was a change in demographics of trips after data cleaning. For casual members, the percentage of users, total miles travelled, and the number of trips were reduced from 92% to 85%, 77% to 63% and 68% to 56%, respectively. For registered members, the percentage of a number of users, total miles travelled, and number of trips increased from 8% to 15%, 32% to 44% and 23% to 37%, respectively. The majority of the trips removed were casual trips.

3.3. Completing The Road Network

The raw data processed using above methods should be matched to the road network in order to get the exact route followed by the cyclists. In contrast to the motor vehicle drivers, the path followed by the cyclists includes those links which may not be present in the base network, such as parking facilities, alleys, or shared use paths. Moreover, most of the problems mentioned above during the map matching of the trips are the result of the absence of these links in the base map. For that purpose, the road network had to be supplemented to predict the path of the cyclists. Most of the added road segments were the alleys, parking spaces, and the parkways, in such a way that all possible links for bicycle travel are included.

4. CONCEPTUAL FRAMEWORK AND METHODOLOGY

4.1. Map Matching

After the base road network is prepared and raw GPS data are processed to exclude the recreational and erroneous trips, the GPS data was matched to the street segments in order to identify all the links that have been traversed during the trip. Although map matching allows us to determine the street segments used by the cyclists, it is difficult to estimate the actual path with high accuracy. The reasons behind this are the inaccuracy of the recorded GPS data points and use of the sidewalks, parking lots and alleys, which are not represented as separate features in the map.

4.1.1. Available Methods

There are vast a number of map matching algorithms that have been developed and used in different studies. Available methods for map matching could be categorized into three groups: geometric map matching, topological map matching, and other advanced techniques (Quddus, Ochieng, & Noland, 2007; Schuessler & Axhausen, 2009).

a. *Geometric Map Matching*

Under this method, the shape of the links of the road network is considered for matching the GPS points based upon the distance of the GPS points to the node/link or shape of the GPS trace. There are in general three methods under geometric Map matching (Hudson, Duthie, Rathod, Larsen, & Meyer, 2012). Geometric map matching can be administered by matching the points to the nearest vertices or nodes of a road segment, which is called point to point matching. Also, the GPS points can be matched to the closest curve in the road network, which is called “point to curve” map matching. In addition to these, the GPS line obtained from the trace of the GPS points can be matched to the road network.

b. Topological Map Matching

This procedure makes the use of geometry of the links as well as the connectivity and continuity of the links. This procedure is based upon the similarity criteria between the trajectory of the GPS points and topological features of the road like turns, curvatures and so on. This offers better results compared to the geometric map matching alone.

c. Advanced Map Matching

In recent years, most of the research use the most advanced approaches for accomplishing the map matching of the GPS points. These processes not only take into account the sequence of the GPS points and the network topology but also considers that, due to the GPS errors or error in network coding, the nearest link might not be the correct one (Schuessler & Axhausen, 2009). Multiple Hypothesis technique is one of the widely used methods under advanced map matching.

4.1.2. Used Procedure

The method used for this study for matching the GPS data obtained from the trip data from the cyclists of bikeshare comes from the study by Hudson et al. (2012), which uses the ArcGIS model for predicting the actual path of the cyclists. This Arc Catalog's Model used by the study was based on the algorithm developed by Dalumpines et al. (2011). The major advantage of using this model was its dependence on the environment of ArcGIS alone. Furthermore, this algorithm successfully implements geometric and topological map matching procedure with the help of Network Analyst function in ArcGIS. The model was built in ArcGIS's Model builder tool based on the comprehensive explanation of the procedure presented by the study by Hudson et al. (2012). Figure 2 is the representation of the result obtained from the implementation of the GIS model.

Following are the steps that are involved in the matching the collected routes depicted by the GPS points to the real road network.

1. Make Route layer from the available network dataset.
2. Load the locations of the each pair of Origin and Destination on this route layer.
3. Create the Buffer of 250 ft. around each GPS points such that all of the buffer for a particular trip will be dissolved.
4. Load the buffered outline for each trip as the barrier to determine the route between the corresponding Origin and Destination.
5. Find the shortest path between the Origin and Destination with the line barrier stated in Step 4. The shortest route created on the route layer (From Step 1) by the Network Analyst in ArcGIS utilizes Dijkstra's shortest-path algorithm.

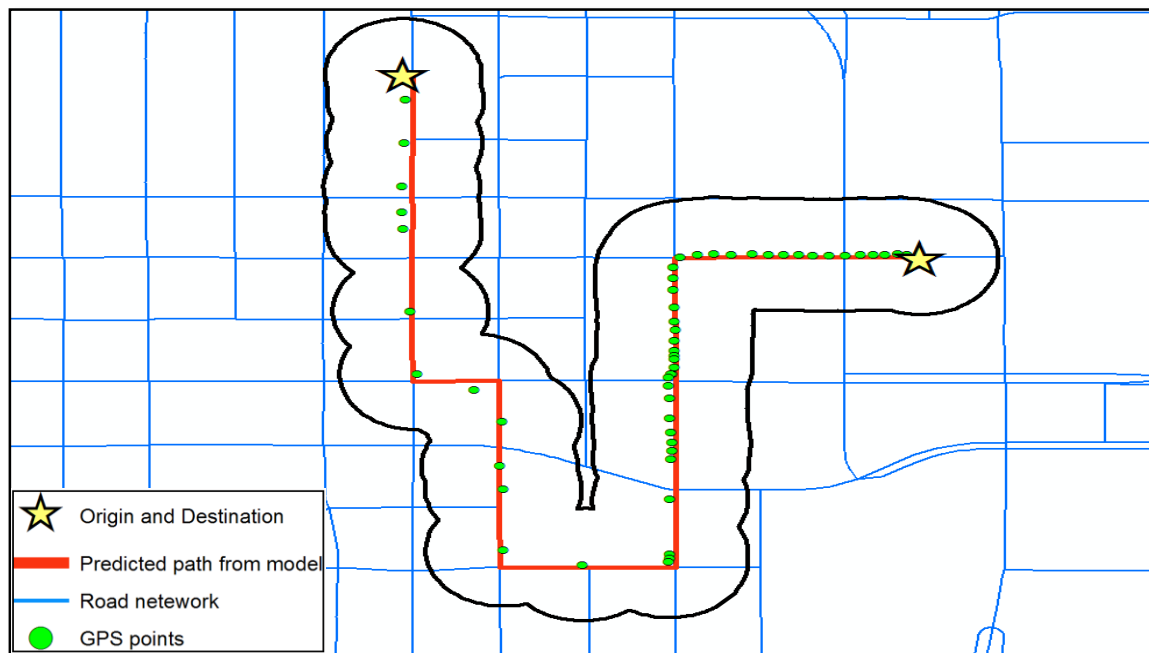


Figure 2 Result of the ArcGIS Model Used for Map Matching of GPS Points

4.1.3. Issues and Solutions

There are several reasons that prevent the algorithm to determine the actual routes. These issues are described under following headings:

Discontinuous/incomplete road network

For the map-matching to yield accurate results, the complete street network having all the segments of the route adopted by the riders is the basic as well as an important prerequisite. Hence, for a successful map matching of GPS points, the road network should be accurate, and should represent the real street network. Otherwise, this could result in failure in accurate representation of road traversed. For instance, for the trip shown in Figure 4(b), the GIS Model cannot determine the actual path of the user because some street segments like alleys, parking areas, parkways, etc. that are used by the cyclists are missing in the available street network. The solution to this issue is the addition of those links to the street segments.

Frequency of the recorded GPS points

The buffer of 250 feet is used for creating the restriction to determine the actual route of bikeshare users. This value of 250 feet is determined based upon trial and error with the sample trips, which provided the highest accuracy of matched trips. For the successful determination of the actual route, the distance between two consecutive points should be less than 500 feet (Euclidean distance). Since the frequency of GPS reading varied from 1 per minute to 25 per minute, some of the points were more than 500 ft. apart. In those cases, separate buffer layers were formed for the same trip. Figure 3 shows the similar case when continuous route could not be determined between origin and destination. This is a limitation of the Map matching procedure used in this study due to sparse GPS points.

Misrepresentation of the trip due to small/large loops inside the trip

Some of the trips might have a loop in it. These kinds of trips are most likely recreational trips. When this model is used for determining the length of these trips, the model will underreport the actual length for this kind of trips. In figure 4(a), the user

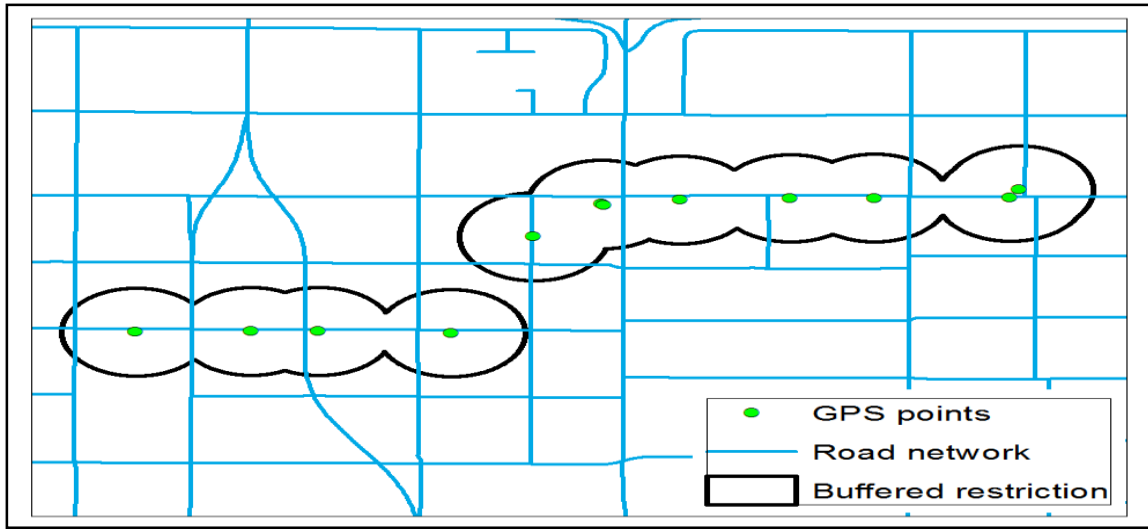


Figure 3 Inability to Merge Buffers due to Sparse Consecutive Points

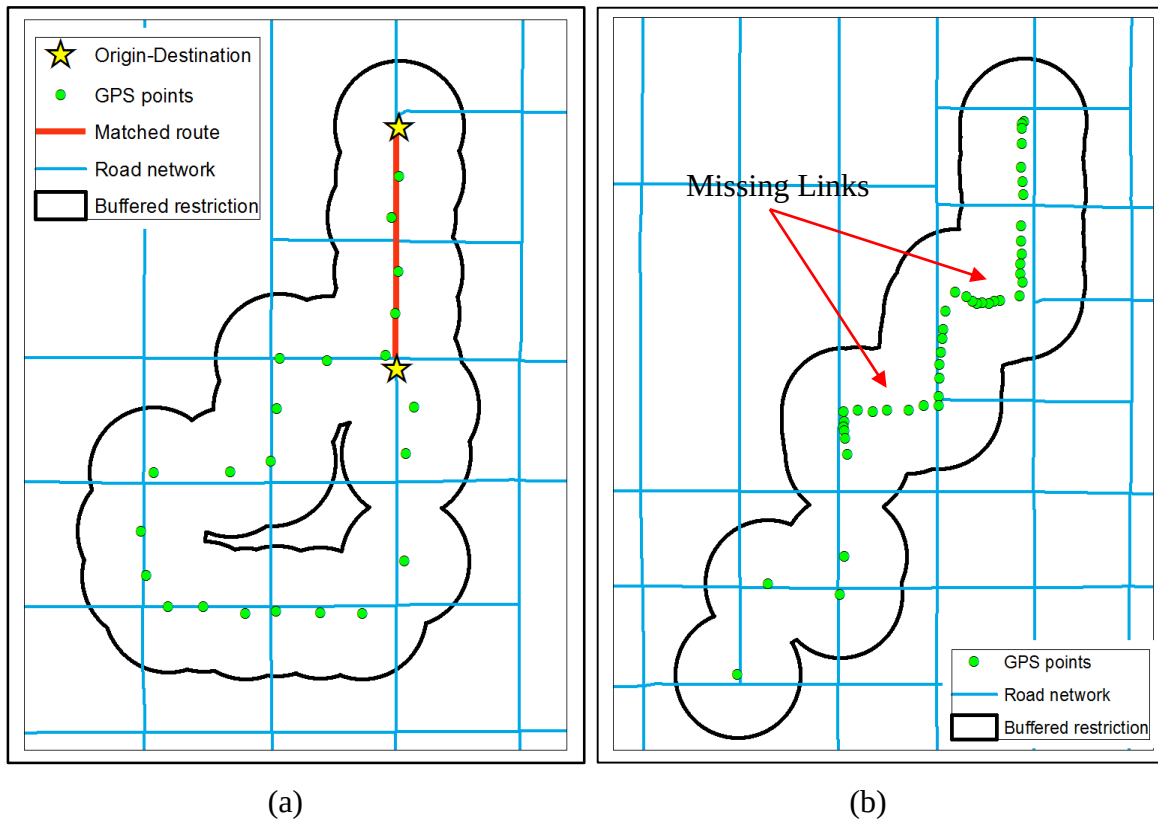


Figure 4 (a) Failure to Predict a Trip with Loops (b) Missing Links in the Trip

had made a tour at the start of the trip and then proceeded to the destination. In this case, the length of the trip determined by the GIS model (shown as a red line) doesn't represent the actual trip. These kinds of trips were removed by using the data cleaning criteria mentioned in Data Cleaning section 3.2.

4.2. Choice Set Generation

To predict the route choice from among the routes that are considered by cyclists, the possible routes for each pair of origin and destination should be identified. Five alternative routes were generated for each Origin and Destination using Network Analyst extension in ArcGIS 10.1. In total, there were six alternatives: five non-chosen and one chosen alternative. Figure 16 in the Appendix shows the set of alternatives for a single trip. The Simple Labelled Route method was used to generate the five non-chosen alternative routes (M Ben-Akiva, Bergman, Daly, & Ramaswamy, 1984). In this method, the shortest path between origin and destination was determined such that certain attributes of the path was either maximized or minimized. The five alternatives were created by maximizing use of bicycle-specific infrastructures along the route and minimizing length, the number of signalized intersections, the proportion of one-way road segments and the number of junctions separately. One way restrictions were not considered while generating these alternative routes to create the feasible alternative routes because some segments of one-way streets might have two-way bike lanes, and even when there is no bike lane many were observed riding wrong way or using the sidewalk.

Characteristics for observed path and five alternative paths generated for the study are as follows:

1. Observed path

This is the actual path followed by the bikeshare users. This is obtained by using the method described in section 4.1.2.

2. Second shortest path

This path minimizes only distance along the route without considering other factors. However, this path was created in such a way that it is independent of the shortest path. This was done because most of the shortest path were exactly same as the observed path. Ten points were created at equidistance along the shortest path, and the second shortest path was created avoiding those points along the route.

3. Path that maximizes the use of bicycle friendly facilities

This path maximizes the use of bicycle friendly infrastructure, i.e., local streets and street with bicycle-specific facilities. This was done with the built-in function of the ArcGIS network analyst, which was capable of imposing the preference on the travel along these segments that fulfilled our definition of bicycle friendly infrastructures.

4. Path that minimizes the number of signals

This path minimizes the number of signalized intersections along the route. This was constructed by imposing the restriction on segments adjacent to the intersection with traffic signals. Since there are a large number of signalized intersections in the area where Grid Bikes Share is located, this is one of the feasible characteristics of the alternative route.

5. Path that minimizes the number of junctions/links

This path minimizes the number of nodes or junctions along the route. First, the field was added in the shapefile of the road with a constant value of 1 for all the links of the roads. Then, shortest path algorithm was utilized with the added field as the impedance i.e. the path was created such that the summation of a new field, which is proxy to a number of links, is minimized. The advantage of this path is that cyclists can avoid conflicts with cross streets and driveways.

6. Path that minimizes the use of one-way street segments

This path minimizes the use of one-way segments. This path was created such that the built-in algorithm of ArcGIS will try to minimize the inclusion of one-way segments as far as possible but would not completely avoid it. This was done by introducing the customs value for one-way street segments to increase the cost of traveling along that route, which makes the segment less attractive while solving for the routes.

4.2.1. *Removing the Identical Alternatives*

The most important characteristics of the Discrete Choice Model is the independence of the alternatives. Since the alternatives used for the analysis were created by employing Simple Labeled Route technique, there were several alternatives for a single O-D pair that were exactly the same. For instance, the route minimizing the distance might be identical to the route minimizing the signalized intersections. In this case, one of the alternatives have to be deleted and the trip would be left with only five alternatives in total. Table 1 shows the result of removing identical alternative routes. This leaves 31% of the O-D pairs with six alternatives while 4% of the O-D pairs are left with two alternatives. Approximately 84% of the trips have at least four alternatives available for the final study.

Table 1 Percentage of Trips with Different Number of Alternatives

Number of Alternatives	Percentage of total trips
TWO	4%
THREE	10%
FOUR	22%
FIVE	33%
SIX	31%

4.2.2. *Calculating the Attributes of Alternatives*

After the map matching, next task was to find the attributes of the each trip. As we have the attributes for all the links along the route, we joined the attributes of the road segment to the segment of the trip traversing it. For the variables like AADT and posted

speed limit, path attributes are generated as the length-weighted sum of each link's value of AADT and speed limit. The basic principle behind this is distributing the attributes to all the paths on the basis of the proportion of the total length of the route.

4.3. Discrete Route Choice Model

Discrete route choice model has been used as a major modeling technique for predicting bicycle route choice as well as mode choice in a number of previous studies (Broach et al., 2012; Sener et al., 2009; Stinson & Bhat, 2003). Discrete choice models empirically model and analyze the decision maker's preferences among the set of alternatives available to them, and these analyses represent the behavior of each individual rather than a group of individuals. For this study, we used the standard random utility maximizing framework. This means that utility of each route is dependent upon attributes of each alternative routes and an unobserved stochastic component. The main basis of the discrete route choice model is that people always choose the alternatives in such a way that overall utility is maximized. For instance, If an alternative i chosen within the choice set of alternative routes C_n , then the utility as perceived by the person n is given by:

$$U_n = \beta x_n + \varepsilon_n$$

Where x_n is the value of the attribute, βx_n is the observed component and ε_n is the unobserved component capturing uncertainty.

Each of the users has to make a choice among the alternatives and decision of selecting this choice is dependent upon the attributes of each alternatives. This can provide the results explaining how cyclists rate different bicycle-specific infrastructures along their route, volume on the road, the number of signalized intersections, stops and so on. Hence, with this model, we can forecast how the behavior of people will change when the attributes of the alternatives are changed.

Based upon the number of available alternatives, discrete choice models can be broadly classified into two categories:

1. *Binomial choice model*: It is used when there are two available alternatives.
2. *Multinomial choice model*: It is used when there are more than two available alternatives.

The most prominent types of discrete choice models are logit, generalized extreme value (GEV), probit and mixed logit (Train, 2009). The Multinomial Logit (MNL) model is used for this study. Multinomial Logit Model is the simplest among the family of logit models. According to this model, the probability of choosing the alternative i among the alternatives available in the set C_n is given by:

$$P(i | C_n) = \frac{\exp(V_{in})}{\sum_{j \in C_n} \exp(V_{jn})}$$

Where,

C_n = the available choice set of alternatives

i = the chosen alternative

j = any alternatives which within C_n

V_{in} = Utility of the chosen alternative i

V_{jn} = Utility of the chosen alternative j

4.3.1. Assumptions of Multinomial Logit Model

The Multinomial Logit Model assumes that the population is homogenous. This means that the behavior of cyclists, while selecting the route among the alternatives, should be similar. The bikeshare dataset was divided into two group of users: casual users and registered users, which provided the homogenous group for analysis as compared to the unsegmented bikeshare users.

For the dataset to be fit for the discrete choice framework, the set of alternatives should fulfill three characteristics. First, the set of alternatives—“choice set”—should be mutually exclusive. This means that MNL assumes independence between the alternative paths. This property of MNL is termed as Independence from Irrelevant Alternatives (IIA). However, to have completely different set of alternative route is not possible as people choose the route based on the attributes of the link but not the attributes of the whole route. If this property is not fulfilled, the MNL model will estimate the high utility of the overlapping paths. The explanation for this is that, although MNL considers overlapping routes as distinct alternatives, for the individual, these are only minor variant to the single alternative (Broach et al., 2012).

Second characteristic is that the alternatives should be finite, which holds in our case. Third, the alternatives should be exhaustive, i.e. it should contain all the feasible alternatives. However, irrelevant alternatives might not be included. Hence, in this study the available alternatives are limited to the feasible set of alternative routes.

4.3.2. Path Size Logit Model

Due to the computational benefit of the simple MNL, modification in the original model was proposed to retain the underlying MNL structure. But, the utility of the overlapping paths is overestimated for MNL. In order to estimate the valid utility functions and predict the choice probabilities correctly, the correction should be made in order to solve the error formulated due to overlap between the alternatives.

There are several methods like C-logit, Path Size Logit (PSL), and Path Size Correction Logit (PSCL) models that could be employed to correct the deterministic part of the utilities by including a commonality specification in the deterministic part. In this study, Path Size Logit model is used, which is relatively simple and performs well compared to other complex models (Bekhor, Ben-Akiva, & Ramming, 2006).

First of all Path Size Factor for the alternative i is calculated, given by following equation (Ramming, 2001)

$$PS_{in} = \sum_{a \in \Gamma_i} \frac{l_a}{L_i} \frac{1}{\sum_{j \in C_n} \left(\frac{L_j}{L_i}\right)^\gamma \delta_{aj}}$$

Where,

l_a = length of the link a

L_i = length of the alternative i

Γ_i = set of the links of alternative i

$\delta_{aj} = 1$ if j includes the link a , 0 otherwise

γ = long-path correction factor, which is considered 0 in our case.

The first term in the summation, l_a/L_i , is a weight by which link-specific terms are summed to form the Path Size. The second term may be thought of as a link size contribution. For a link used by only one path, this term is equal to one, so that path accrues the full-size contribution from that link. That is, the total path size also depends on the link size contributions accrued from other links in the path. When more than one path shares a link, the “link size” of one is split equally among the paths.

The long-path correction factor (γ) is used as a positive scaling term to penalize the very long routes among the alternative routes. As there are not very long alternatives in our choice set C_n due to the less number of alternatives, we will use its value as zero which will give us the Path Size Logit (PSL) model (Moshe Ben-Akiva & Bierlaire, 1999). After the correction factor of PSL, the resulting probability that the alternative i is chosen from choice set C_n is given by

$$P(i | C_n) = \frac{\exp(V_{in} + \ln(PS_{in}))}{\sum_{j \in C_n} \exp(V_{jn} + \ln(PS_{jn}))}$$

Where, PS_{in} will have values between 0 and 1, and hence, the $\ln(PS_{in})$ is always negative. This implies that the utility decreases when there is more overlap between the

alternatives, since by introducing path size factor, we are introducing the penalty for the route. The general form of the deterministic part of the utility function will be as follows:

$$U_n = \beta x_n + \beta_{PS} * \ln(PS)$$

This model was estimated through the freely available software Easy Logit Modeler (Easy Logit Modeler).

4.3.3. Distance Trade-off Calculation

To aid in interpretation, we can estimate marginal rates of substitution between distance and other explanatory variables. The distance trade-off for a unit change in attributes can be determined after estimating the utility coefficients of the attributes from following equation for the non-unit changes:

$$Equivalent \% \Delta distance = \left(\exp \left(\Delta attribute * \frac{\beta_{attributes}}{\beta_{\ln(distance)}} \right) - 1 \right) * 100$$

Where β is the coefficient of the attributes of the path estimated from the model.

4.4. Variables Utilized in the Model and Hypothesis of Utility

The variables used for the estimation of discrete route choice model, which are collected based upon the previous researches and current availability of the data, are described in Table 2 with the corresponding hypothesized direction of utility.

The variables are transformed for the rational interpretation of the results obtained from the model. Natural log of the length performed better in the model than length alone. Similarly, introducing length of the bike-specific facilities and one-way road segments in a model will give a wrong interpretation of results. So, the presence of these attributes should be expressed as the proportion of total route. Therefore, proportion of bike-specific facilities and number of left turns, right turns and signalized intersections per mile facilitate the calculation and interpretation of the distance trade-off.

Table 2 Variables Used for the Choice Model and Corresponding Hypothesis

Original Variables	Transformed Variable	Description	Hypothesis for Utility
Length (miles)	Ln(Length)	The length of the actual route obtained from GPS points and alternative routes created from ArcGIS.	Negative Utility for longer route
AADT (vehicles per day)	AADT/1000	Annual Average Daily Traffic for a segment of road	Negative Utility for high AADT
Posted Speed Limit (Mph)	(Unchanged)	Maximum allowable speed on a road segment	Negative Utility for high-speed roads
Left and Right turns	Left turns per mile Right turns per mile	Total number of left or right turns made in the routes	Negative Utility for both type of turns
Number of Signalized intersections	Number of signalized intersections per mile	Total number of signalized intersections traversed along the route	Negative Utility for higher number of signalized intersections
Number of BMV crashes	Number of bicycle motor vehicle crashes per mile	Total number of Bicycle Motor Vehicle crashes from 2010 to 2013	Negative Utility for higher number of crashes per mile
Length of bike-specific facilities (mile)	Proportion of bike-specific facilities	Total length of the road segments having bike- infrastructures along the route	Positive Utility for higher proportion of bike facilities
Length of one-way segments (mile)	Proportion of one-way segments	Total length of the road segments with one-way restriction along the route	Negative Utility for higher proportion of one-way segments
Path Size	Ln(Path Size)	Correction factor for the error due to commonality of the alternative routes	Positive or Negative Utility based upon the alternatives
Peak Hours	Dummy Variable	Time period between 7 am to 9 am and 4 pm to 6 pm	More sensitive to other factors on peak hours

5. ANALYSIS AND RESULTS

5.1. Demographics of the Bikeshare Users

In this analysis, all those users having membership type of student annual, monthly and annual are named as registered subscribers as they resemble the group that commits use of bikeshare for the longer time interval. Conversely, those users who pay a per-trip fee before riding bikeshare are named as casual subscribers.

The registered members comprise approximately 15 % of the 1,866 users but account for 37 % of the total 10,476 miles traveled. The results summarized in Figure 5 show a high proportion of the number of trips made by the registered members. After cleaning the data (i.e., removing recreational tours and erroneous trips), the final dataset was reduced to 9,101 observations of which 43.5 % of the trips were made by registered members and 56.5 % of the trips made by casual users.

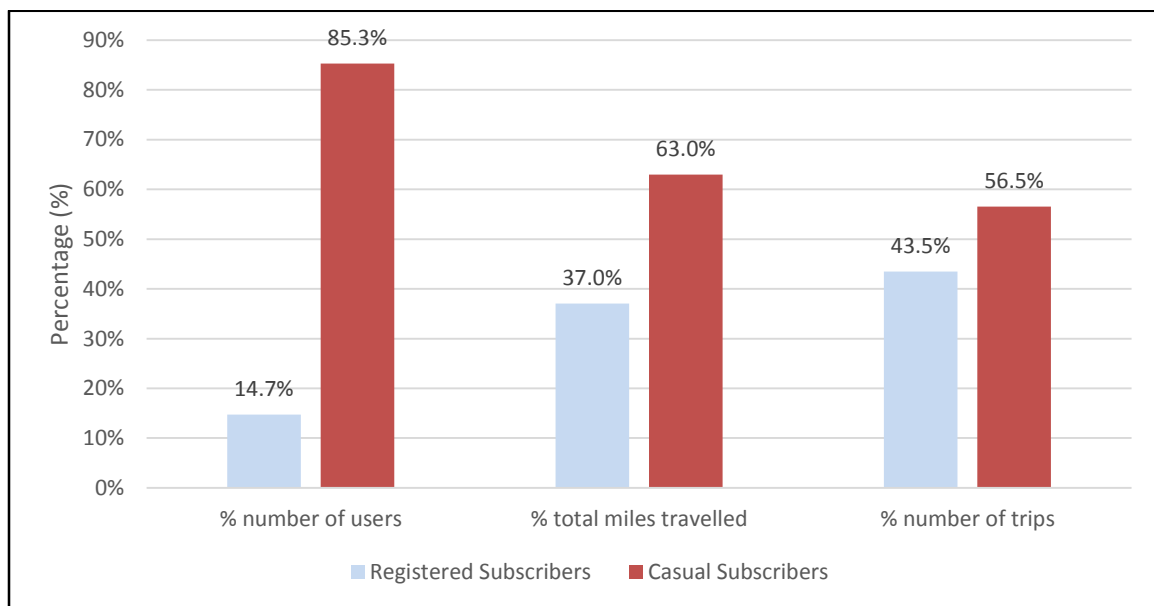


Figure 5 Distribution of Registered and Casual Subscribers

Figure 6 shows that trips by casual users increase steadily from the morning and peak at 5 pm and then drop off into the night. However, trips by registered users peak at 8 a.m., 12 p.m., and 5 p.m. Figure 7 shows that most of the casual trips are made during the weekend, with a number of weekend trips being approximately equal for the weekdays. There is minor variation in activity of registered users over the week.

The trip behavior of the two user groups differed. The mean distance of the trips for registered and casual users were 1.0 (std. dev: 0.64) and 1.3 (std. dev: 0.95) miles, respectively; and similarly, the mean duration of the trip was 9.5 (std. dev: 7.2) and 14.5 (std. dev: 11.7) minutes, respectively. Registered members were making high percentage (69%) of trips less than 10 minutes of travel time. In contrast, 55% of casual user's trips are more than 10 minutes. Similarly, only 2% of the registered user's trips and 10% of the casual user's trips have travel time greater than 30 minutes.

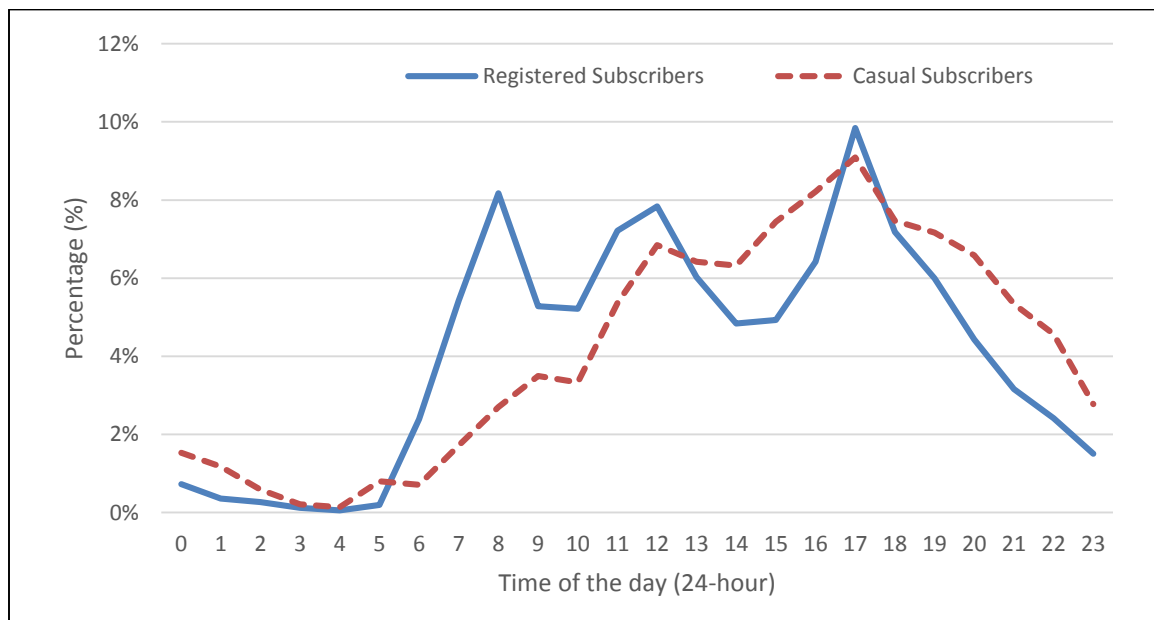


Figure 6 Time of the Day Variation of Percentage of Trips for the Bikeshare Users

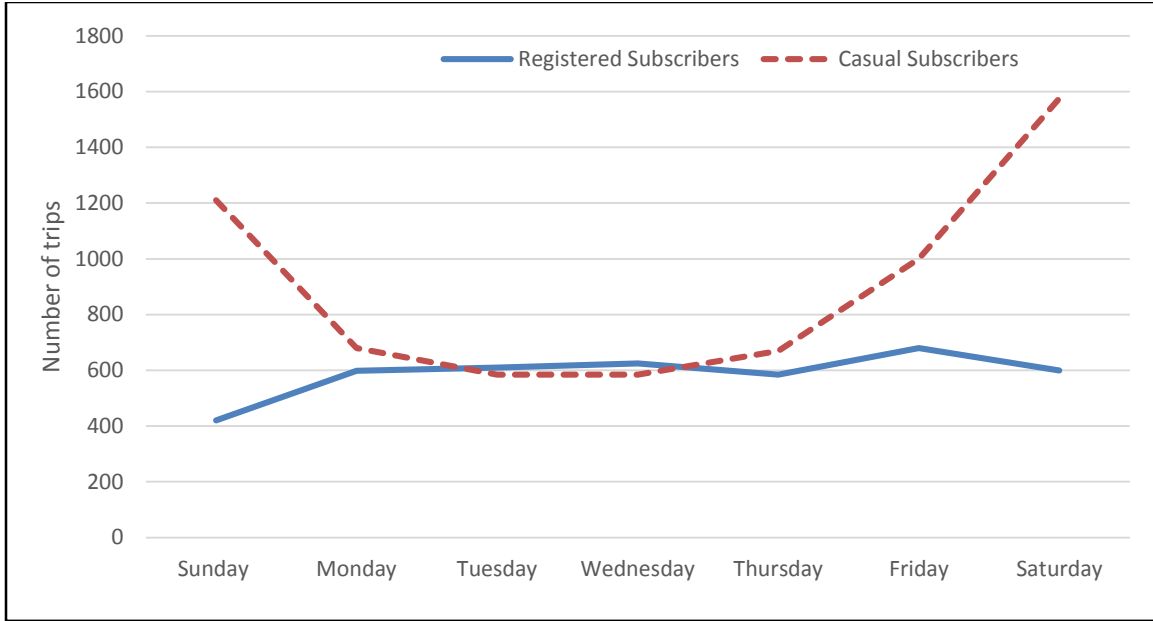


Figure 7 Day of the Week Variation of Number of Trips for the Bikeshare Users

5.2. The Wrong Direction Riding

Riding on the wrong way or on a wrong side of the road is one of the most common and potentially dangerous behaviors among cyclists. Wrong direction riding not only causes crashes but also induces the negative attitudes of drivers towards the cyclists. The advantages of using wrong directions, as perceived by the cyclists, are numerous. For instance, this behavior might reduce the number of crossings along their route and reduce the total trip distance. Some people might feel that traffic coming towards them is much safer. Sometimes, poor street design like long median divider might be restricting riders to make the required turn and obliges them to ride on wrong direction of a road.

For the evaluation of wrong direction riding behavior of bikeshare users, trips on forty different street segments were chosen from the Downtown of Phoenix as shown in Figure 15 on Appendix. This is the area where most of the stations are located, and contains highly traveled road segments. The comparison of wrong direction riding

behavior was done between registered and casual subscribers for one-way streets vs. two-way streets and streets with bike facilities vs. streets without bike facilities.

With the GPS data available in the study, it is almost impossible to track whether they are using the sidewalk or the main road for their travel. In this study, the road is divided into two directions using centerline of the road, and the directions of the road are compared with the direction of the trip made by cyclists to determine the proportion of trips being made on the wrong direction of the road. Link-level violation was determined for forty segments.

Wrong direction riding is ascertained based upon the first and last GPS point on either side of the road. For instance, if a cyclist is riding in the right direction of a road for half the length of the two-way road segment and crosses the centerline to ride on the wrong direction along the same road segment, the trip is counted as wrong direction riding and right direction riding separately because both directions of the road are analyzed individually. Paired Sample *t*-test was done to ascertain the statistical difference in proportion between registered and casual users riding against the flow of traffic.

Figure 8 summarizes the proportion of trips being made on the road against the traffic for casual subscribers and registered subscribers on one-way and two-way roads. In both the cases, casual members are more involved in wrong direction riding behavior. The percentage of trips on the wrong direction of the one-way streets is 23% for registered users (28% for casual users). Similarly, the percentage of trips on the wrong direction of two-way streets 20% for registered users (29% for casual users). The difference in the proportion of trips on the wrong direction of the road between two users is significant at 95% Confidence Interval for one-way roads ($p=0.025$) and two-way roads ($p=0.027$).

Figure 9 presents the proportion trips on the wrong direction of the road segments with or without bike-specific facilities. This also shows the lower violation rate among

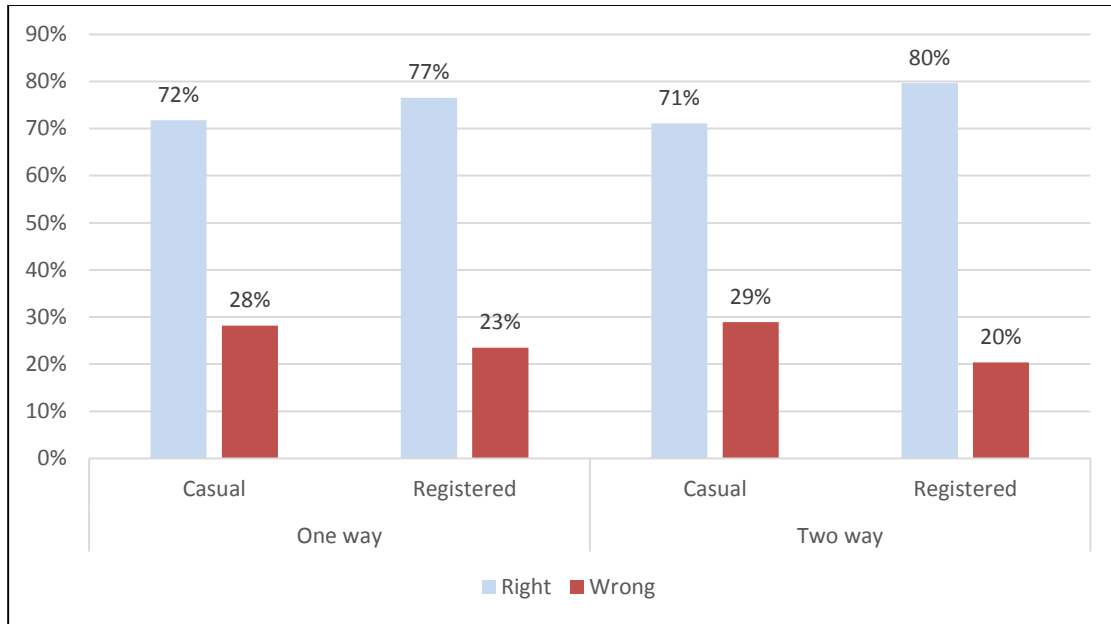


Figure 8 Wrong Direction Riding for One-way and Two-Way Road Segments

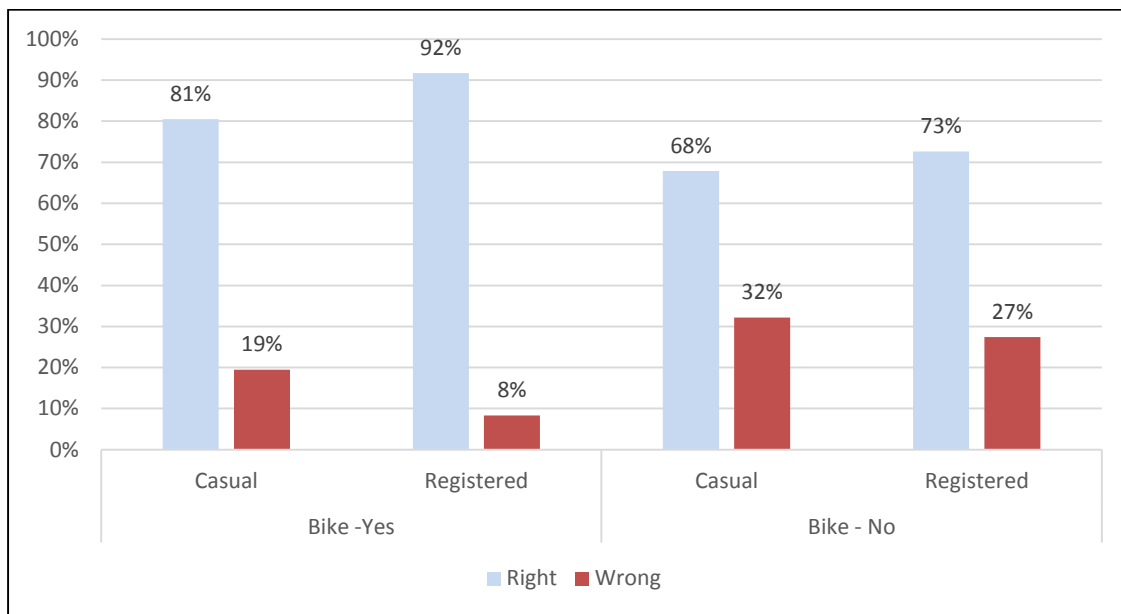


Figure 9 Wrong Direction Riding on the Segments With or Without Bike Facilities

registered users relative to the casual users. The percentage of trips in the wrong direction with bike facilities is 8% for registered users (19% for casual users). Similarly, the percentage of trips on the wrong direction of the streets without bicycle facilities 27% for registered users (32% for casual users). Difference in proportion of trips on the wrong direction between two users is significant at 99% Confidence Interval and 90% Confidence Interval for roads with bike facilities ($p=0.002$), and road without bike facilities ($p=0.095$).

5.3. Facility Usage Assessment

Once the raw data was cleaned, processed data was matched to the base network and final a map was obtained illustrating the number of trips for each road segments. Volume of the bicycle trips made on the streets of Downtown of Phoenix are illustrated in Figure 17 in Appendix.

There is a high number of bicycle trips along the Central Ave and 1st Ave, both of these roads connect to the center of the downtown of Phoenix. Bikeshare stations in the center of downtown are among highly used stations. High number of population, land use pattern, concentration of bikeshare stations and requirement for improved accessibility (especially for a short distance) are the major reasons for a high density of bicycle trips on the roads of central downtown compared to roads outside of the downtown.

5.3.1. *Distribution of Different AADT Levels or Speed Limits on the Observed Route*

Figure 10 and Figure 11 illustrates how the different categories of users are using the roads based upon the AADT and posted speed limit on the road respectively. Highest proportion (56.4%) of registered users are using the roads with AADT less than 5000 vehicles per day. But, casual users are riding more on the roads with high volume as compared to registered users. Figure 11 indicates that approximately 45% of the trips are made on the roads with a speed limit less than 30 mph.

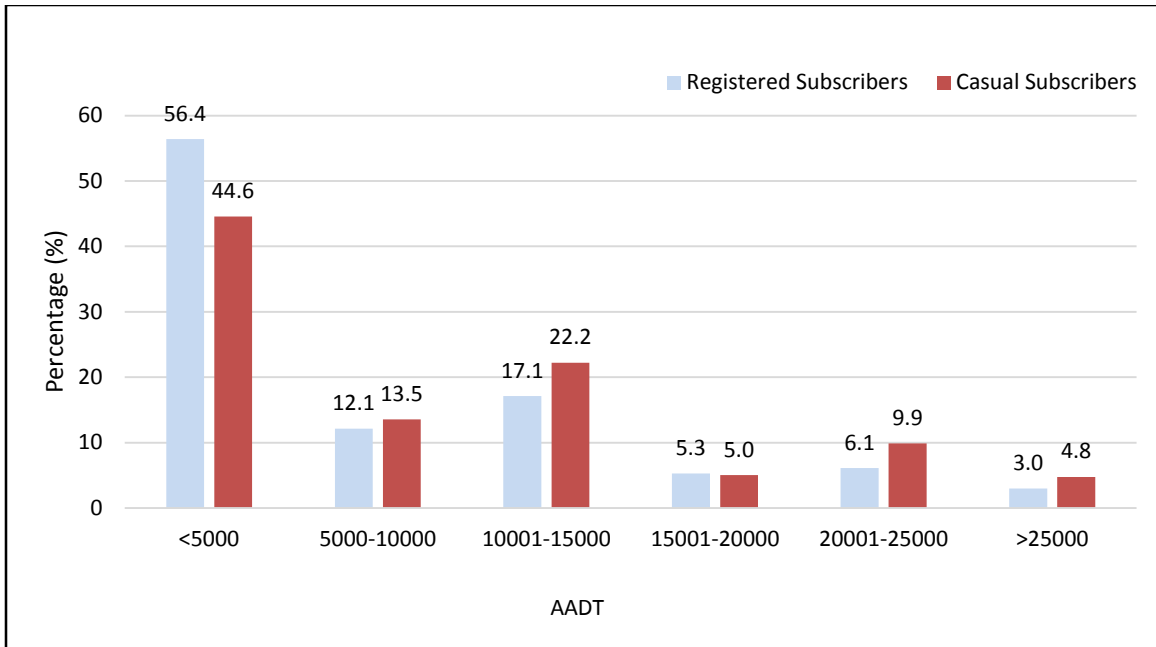


Figure 10 Distribution of Travel Distance Over Range of AADT

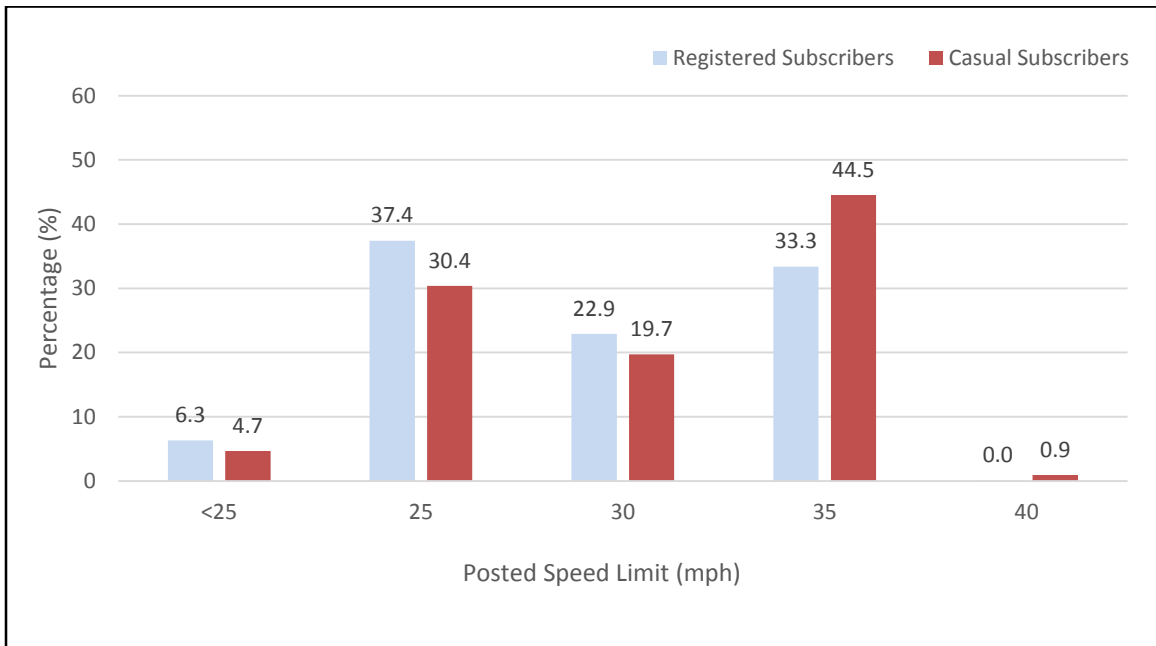


Figure 11 Distribution of Travel Distance Over Range of Speed Limit

5.3.2. *Distribution of Different Roadway Infrastructures on the Observed Route*

Figure 12 and Figure 13 illustrates the usage of different bicycle-specific facilities and roadway infrastructures respectively. The statistics show that there exists very less difference between the use of the bicycle specific infrastructures between registered and casual users. The proportion on the use of bike lane is slightly higher for casual users. On the other hand, local roads are most favorite roads among registered users. There is the difference in 10% between uses of local roads among two groups of bikeshare users. Registered users make slightly higher travel on parking space and parkways while casual users travel slightly more on trails.

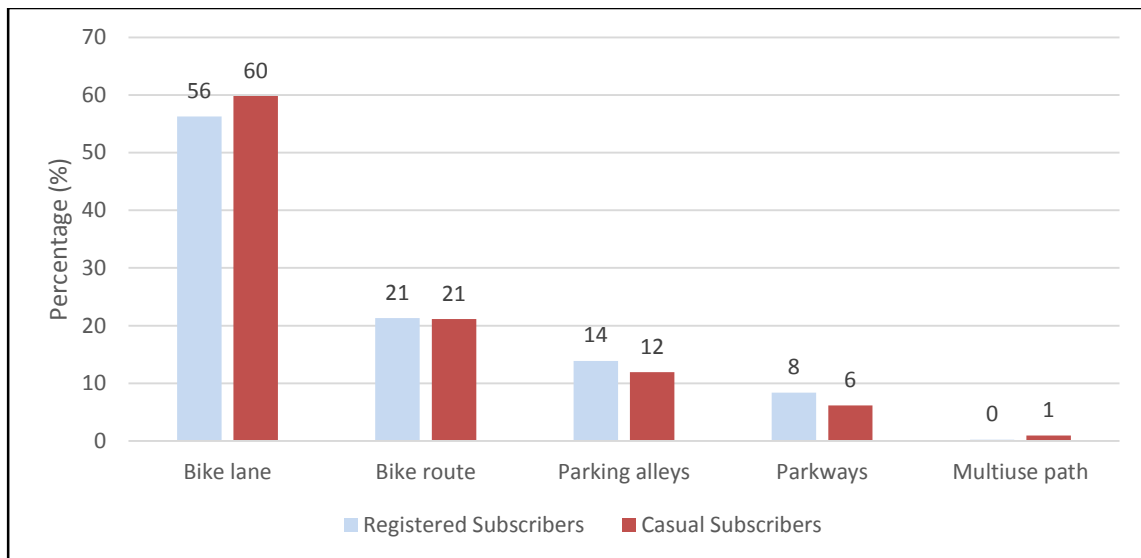


Figure 12 Usage of Different Bicycle Facilities



Figure 13 Usage of Different Roadway Infrastructures

5.4.Route Choice Assessment

The second objective of the study is to find the route preference of the bikeshare users, which is done by modeling the route attributes of the chosen attributes against the generated alternative routes. Before modeling, the correlation between variables was analyzed. The speed limit was found to be highly correlated with AADT along the route (Correlation Coefficient=0.60, p-value=0.000). AADT was included and the speed limit was excluded from a final model based upon likelihood ratio test. AADT was scaled to a smaller value dividing by 1000 to obtain more appropriately scaled parameter estimates.

Furthermore, bicycle-vehicle-crash per mile was tested in the model because it was hypothesized that number of crashes would act as a proxy to the dangerous road. However, this variable was insignificant. Bikeshare users do not likely know the number of bicycle-motor vehicle collision (or risk proxy) and could not respond accordingly to avoid the routes with a high number of crashes. Similarly, effect of day of week on the route choice behavior is not included in the final model because it was found to be insignificant.

Table 3 presents results for the estimation of the Path Size Logit route choice model. Two models were developed, one for registered subscribers and one for casual subscribers. Using the value of coefficients obtained from the route choice model, the distance value of the unit change in the value of attributes are derived and presented in Table 4. Length, the number of left and right turns, the proportion of one-way segments and traffic volume provides negative utility for a route.

The presence of bike-specific infrastructures and signalized intersections have positive utility while choosing the route. Effect of length of the trip and proportion of bike-specific infrastructures have positive utility as compared to the trip made on off-peak hours. However, the presence of one-way road segments had the negative utility for route selection in peak hours as compared to off-peak hours' rides. Interaction of the dummy variable with the other variables in the model is found to exert no significant on route choice decisions made by the riders.

Table 3 Estimation of Utility Coefficients

Variable	Registered Subscribers			Casual Subscribers		
	Est. Coeff.	t-stat	Rand. Err.	Est.Coeff	t-stat	Rand. Err.
ln(length)	-4.64	-17.77	0.261	-2.83	-15.05	0.188
Proportion of bike facilities	2.77	17.78	0.156	2.16	19.32	0.112
Number of left turns per mile	-0.14	-10.22	0.014	-0.17	-13.61	0.013
Number of right turns per mile	-0.14	-9.91	0.014	-0.13	-10.66	0.012
Proportion of one way	-0.43	-3.63	0.119	0.11	1.20**	0.090
Numbers of signals per mile	0.25	17.99	0.014	0.24	20.81	0.012
AADT/1000	-0.16	-21.27	0.007	-0.08	-15.80	0.005
ln(length)*Peak hour	-3.97	-6.63	0.598	-2.45	-5.19	0.472
Proportion of bike * Peak hour	0.93	3.23	0.288	0.46	1.92*	0.239
Proportion of one way * peak hour	-0.57	-2.61	0.219	-0.49	-2.42	0.202
ln(PS)	1.26	15.77	0.080	1.17	18.82	0.062
Log Likelihood at Zero		-5587.23			-7533.47	
Log Likelihood at Convergence		-4140.59			-6284.41	
Adjusted Rho Squared w.r.t. Zero		0.2569			0.1643	
Number of Cases		3958			5143	

** - Insignificant

* - Significant at 90% Confidence Limit

Table 4 Distance Value (%) for Unit Change in Attribute

Variable	Distance Tradeoff (% distance)	
	Registered Subscribers	Casual Subscribers
Proportion of bike facilities	-44.9	-53.3
Number of left turns per mile	3.2	6.3
Number of right turns per mile	3.1	4.8
Proportion of one-way	9.8	-3.7
Number of signals per mile	-5.2	-8.2
<i>Peak hour (Baseline: Off peak hour)</i>		
Proportion of bike facilities	-20.9	-17.1
Proportion of one-way facilities	15.5	22.0

6. DISCUSSION

Results indicate that the average length of travel made by registered users are less than that of casual users, and number of trips made by registered users are highest at peak hours. This suggests the commute nature of the trips. Since they registered for the bikeshare with a long-term subscription, they might be using it often for the utilitarian purpose. Additionally, their preference towards a low-volume and low-speed road strengthens the fact that they are generally local cyclists who know more about the alternative roads in the area and choose the roads which are easier and safer to travel. Registered users are less likely to go against the flow of road during their travel as compared to casual users. Furthermore, less proportion of wrong direction riding for both groups in the roads with bicycle facilities highlights the importance of bike-specific facilities to reduce the behavior of riding against the traffic. Hence, it can be concluded that registered users are making a shorter and safer ride than casual users.

Table 3 presents results for the estimation of the final Path Size Logit route choice model. Two models were developed, one for registered subscribers and one for casual subscribers. The negative coefficient for distance variable supports the well-known fact that cyclists prefer shorter routes among available alternatives unless there are other desirable attributes on other alternatives that outweigh the advantage of short distance. The magnitude of the coefficient suggests that registered users are more sensitive to the length of selected route compared to casual users. This is likely because registered users use bikeshare to make utilitarian trips in most cases, which is reinforced by the time of the day and week of the day variation is shown in Figure 6 and Figure 7. The average length of the observed path for registered users and casual users is 6.9% and 8.3% higher than the average length of the shortest path, respectively. This statistics bolsters the difference in preference of these two groups over the length of the route but also points to other factors influencing route choice.

AADT is associated with negative utility for both categories of users. The risks associated with travel along high volume roads, which affects the perceived safety of the users, is likely a major reason for the disutility towards high volume roads. The extent of disutility is slightly higher for registered users. Registered users are those users who are committed to using the bikeshare system, and subscribed for a month or year. Hence, they are most likely to have information on which roads have high volumes of vehicles in the surrounding network and avoid them as far as possible. AADT was interacted with the peak hour (7 am-9 am and 4 pm-6 pm) to test if the time of the day has an effect on route choice. However, the disutility of high AADT perceived by bikeshare users does not change significantly over the time of the day.

Both groups of users have high preference towards including bike-specific facilities in their route. This finding supports previous literature that asserts the preference of bike lanes and bike routes among cyclists. Use of bike lanes, shared paths or multiuse paths has many inherent advantages such as separation from high-speed traffic and increase in perceived safety and freedom to ride at their preferred speed. Travel on the bike-specific facilities is equivalent to decreasing distance by 44.9% (53.3% for casual users).

Registered users avoid including one-way road segments in their trip as far as possible. The route choice behavior of casual users is not definitive regarding the inclusion of one-way road segments. Registered users are aware of the information on the competing routes, which allows them to choose a route that minimizes or avoid one-way street segments. On the other hand, casual users (sometime tourists and potentially infrequent cyclists) are unaware of the alternatives and could not avoid these segments.

A number of signalized intersections along the route is a positive factor for the selection of the route. The coefficients for the number of signals per mile are equal for both groups of bikeshare users. This result could be counterintuitive in that the signalized intersection on the route decreases the utility of the route because these intersections add

delay and potential risk. However, this might not be always true. Signalized intersections provide relatively safe, protected crossings of large roadways. It is also reasonable that, in a grid network, cyclists tend to ride on main arterials and tend to avoid routes that require them to cross un-signalized (e.g., midblock) minor street crossings. Furthermore, the downtown Phoenix area is highly signalized, making it difficult to avoid signals along various routes. This might explain the positive parameter estimate for the route with signalized intersections.

The number of turns along the route is another factor that cyclists account for while choosing a route. Both, registered and casual users, valued routes with fewer left- and right-turns, but in a different manner as suggested from Table 3. The difference in the value of coefficient shows that cyclists, in general, have a greater aversion to left-turns compared to right-turns, as expected. The higher delay associated with left turns, at signalized as well as un-signalized intersections, and additional safety risk associated left turns compared to right turns could be the main reasons for disutility of this variable for both users. The difference is more pronounced for casual cyclists, but registered users do not seem to differentiate much between left- and right-turns in terms of utility.

Previously, we found that registered users mostly travel on low-volume and low-speed roads. Since left- and right-turns have a similar effect in terms of delay time or difficulty in maneuvering in low volume and low-speed roads mostly traversed by registered users, they might give equal priority to the left and right turns. For avoiding each left turn in a mile, registered users would choose routes that were 3.2% longer (6.3% for casual users). Corresponding additional percentage of route length is 3.1 % and 4.8 % for each additional right turn. This clarifies the comparison between left and right turns on the route made by registered and casual and users.

Effect of time of the day and day of the week was analyzed. For this, morning and evening peak hour (7 am to 9 am and 4 pm to 6 pm) are categorized as peak hour. This variable is interacted with other variables of the model. The proportion of bike-specific

facilities, the proportion of one-way road segments and trip length had significantly different effects compared to off-peak hours. Both users were more likely to make shorter trip avoiding one-way segments and including bike specific facilities. Other variables did not have distinctly different effects at different times of the day.

Travel on bike-specific facilities in peak hour is equivalent to decreasing distance by 20.9% for registered users (17.1% for casual users) compared to off-peak hours. Peak hour traffic and delays associated with travel could be the major reason in selecting shorter routes by both users. This result also bolsters the sensitivity of registered users towards shorter trip length compared to casual users. Similarly, the advantage of fast travel without the interference of the large number of high-speed vehicles during peak hour, in addition to increased perceived safety, can explain the preference of bike specific facilities during peak hour as compared to an off-peak hour. A key unobserved factor that likely varies between times of day, demographics, likely affects the differences as well.

As the utility of overlapping paths is overestimated in the MNL, Path Size Correction is introduced to adjust the utilities for overlap. Since the value of PS lies between 0 and 1 (1 for the unique route), $\ln(PS)$ is always negative. The estimate of $\ln(PS)$ should be positive and significantly different from 1, which is similar to our results (29). This has a meaningful interpretation in the case of the route choice model because this is used to correct the commonality or correlation between the alternatives mentioned in section 4.3.2. The positive value of the coefficient associated with $\ln(PS)$ shows that the correction on the utility of the route will be negative, i.e. the utility of the route will decrease to the extent based upon the extent of overlap.

The findings from this paper can be compared across other studies of cyclists' route choice. Several consistencies exist between this and previous studies that examine conventional cyclists. Negative utility for the longer trips is consistent with all of the previous studies. Length of the route results in negative utility for all of the studies. Consistent positive preferences towards bike facilities can be seen in previous studies

(Broach et al., 2012; Hood et al., 2011; Menghini et al., 2010). A study in San Francisco, California found cyclists are willing to add a mile on bike lanes in exchange of 0.5 miles of ordinary roads (Hood et al., 2011). This result is very near to the result of our study, which indicates that if there is one mile of road with bicycle facilities it is equivalent to 0.55 miles of normal road (0.47 miles for casual users).

A number of turns per mile was found to have distance trade-off value of 4.2% (commute trips) and 7.4% (non-commute trips respectively (Broach et al., 2012). Left and right turns are analyzed separately in this paper. For registered and casual users, distance trade-off is 3.2% for left-turn (3.2% for right turn) and 6.3% for left-turn (4.8% for right-turn) respectively. This gives a ground for comparability of registered and casual users with the commute and non-commute trips. Other studies, however, estimated 17 % distance value for one turn (Hood et al., 2011). This is significantly higher than the value estimated by this study.

Signalized intersections were found to be used while crossing major roadways and turning, traffic volume has a consistently negative impact on route choice (Sener et al., 2009; Menghini et al., 2010), especially for left turns in cross street with high AADT relative to right turns with the same AADT in cross streets (Broach et al., 2012). This result of negative utility for AADT is consistent with this study too. In spite of various studies that identify the possible risk factors for the cyclists making turns or making through movements at signalized intersections (Wachtel & Lewiston, 1994), our study found preferences towards signalized intersections. This might be attributed to either different behavior of bikeshare users or benefits of protected phases through signalized intersections.

7. CONCLUSIONS, LIMITATIONS, AND FUTURE RESEARCH

The main objective of the study was to use the real-time GPS data to determine the route choice preference for bikeshare users. Although, the number of bikeshare system has increased sharply in North America, thorough research on the behavior and route selection is unknown, especially for bikeshare users. This research is unique because it is among the first that looks at route level analysis of bikeshare trip (enabled by GPS), it relies on a system without strict station origin and destinations, and it is operated in a city with a grid transportation system enabling many feasible alternative routes. The results of this study clearly showed the road usage behavior and preferences of various factors in selecting the route for both categories of bikeshare users in the road.

However, there are some limitations to this study which are described in following paragraphs. First, although we have six routes under choice set for the analysis, we do not know about the actual routes that were considered by the bikeshare users while selecting the route and we do not know if bikeshare users have perfect information on all routes, an assumption of revealed preference choice modeling. This is a limitation of all the studies that relies only on revealed preference data. Second, lack of demographic information for the users is another limitation of the data. We do not have any personally identifying information of the users, including some demographic factors, such as age, sex, occupation, income, cycling frequency etc. that could influence route choice of cyclists. A recent result from Capital Bikeshare member survey report has identified that bikeshare users tend to be young, more affluent, white, and male (2014 Capital Bikeshare Member Survey Report). Third, the data is not representative of many urban cyclists and suffers from self-selection bias. But, significant impact on the road can be seen by the approximately 50,000 miles of travel (until the end of July 2015) made on 500 bikes on the city. Hence, this sample of users could not be neglected.

Since most of the origins and destinations were fixed, though this bikeshare system allows non-station origins and destinations, a future study could focus on the influence of the placement of these stations on the route choice model to balance placing stations on visible, busy streets that force users to ride on those streets for station access. The riders tended to value travel distance more than other factors and planners should focus on providing better alternative route information, especially to non-subscribing users, identifying station locations that allow direct access to bike-friendly routes, and improving the safety and operations of routes in the service area in regard to cycling (e.g., lowering speed limits on main corridors).

LIST OF REFERENCES

- Bekhor, S., Ben-Akiva, M. E., & Ramming, M. S. (2006). Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, 144(1), 235-247.
- Ben-Akiva, M., Bergman, M., Daly, A. J., & Ramaswamy, R. (1984). *Modeling inter-urban route choice behaviour*. Paper presented at the Proceedings of the 9th International Symposium on Transportation and Traffic Theory, VNU Press, Utrecht.
- Ben-Akiva, M., & Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions *Handbook of transportation science* (pp. 5-33): Springer.
- Bikesharing.com. The Bike-Sharing World Map. from https://www.google.com/maps/d/u/0/viewer?mid=zGPlSU9zZvZw.kmqv_ul1MfkI&hl=en
- Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, 46(10), 1730-1740.
- Broach, J., Gliebe, J., & Dill, J. (2009). Development of a Multi-class bicyclist route choice model using revealed preference data.
- Buck, D., Buehler, R., Happ, P., Rawls, B., Chung, P., & Borecki, N. (2013). Are Bikeshare Users Different from Regular Cyclists? A First Look at Short-Term Users, Annual Members, and Area Cyclists in the Washington, DC, Region. *Transportation Research Record: Journal of the Transportation Research Board*(2387), 112-119.
- Dalumpines, R., & Scott, D. M. (2011). GIS-based map-matching: Development and demonstration of a postprocessing map-matching algorithm for transportation research *Advancing geoinformation science for a changing world* (pp. 101-120): Springer.
- Dill, J., & Gliebe, J. (2008). Understanding and measuring bicycling behavior: A focus on travel time and route choice.
- Duthie, J., Brady, J., Mills, A., & Machemehl, R. (2010). Effects of on-street bicycle facility configuration on bicyclist and motorist behavior. *Transportation Research Record: Journal of the Transportation Research Board*(2190), 37-44.
- Easy Logit Modeler. from <http://elm.newman.me/downloads>

- Ehrgott, M., Wang, J. Y., Raith, A., & Van Houtte, C. (2012). A bi-objective cyclist route choice model. *Transportation Research Part A: Policy and Practice*, 46(4), 652-663.
- Fishman, E., Washington, S., & Haworth, N. (2013). Bike share: a synthesis of the literature. *Transport Reviews*, 33(2), 148-165.
- Fishman, E., Washington, S., & Haworth, N. (2014). Bike share's impact on car use: evidence from the United States, Great Britain, and Australia. *Transportation research part D: transport and environment*, 31, 13-20.
- Hood, J., Sall, E., & Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. *Transportation letters*, 3(1), 63-75.
- Hudson, J. G., Duthie, J. C., Rathod, Y. K., Larsen, K. A., & Meyer, J. L. (2012). Using smartphones to collect bicycle travel data in Texas.
- Johnson, M., Newstead, S., Charlton, J., & Oxley, J. (2011). Riding through red lights: The rate, characteristics and risk factors of non-compliant urban commuter cyclists. *Accident Analysis & Prevention*, 43(1), 323-328.
- Klasnja, P., & Pratt, W. (2012). Healthcare in the pocket: mapping the space of mobile-phone health interventions. *Journal of biomedical informatics*, 45(1), 184-198.
- Langford, B. C., Chen, J., & Cherry, C. R. (2015). Risky riding: naturalistic methods comparing safety behavior from conventional bicycle riders and electric bike riders. *Accident Analysis & Prevention*, 82, 220-226.
- Larsen, J., & El-Geneidy, A. (2011). A travel behavior analysis of urban cycling facilities in Montréal, Canada. *Transportation research part D: transport and environment*, 16(2), 172-177.
- Menghini, G., Carrasco, N., Schüssler, N., & Axhausen, K. W. (2010). Route choice of cyclists in Zurich. *Transportation Research Part A: Policy and Practice*, 44(9), 754-765.
- Quddus, M. A., Ochieng, W. Y., & Noland, R. B. (2007). Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transportation Research Part C: Emerging Technologies*, 15(5), 312-328.
- Ramming, M. S. (2001). *Network knowledge and route choice*. Massachusetts Institute of Technology.

- Räsänen, M., & Summala, H. (1998). Attention and expectation problems in bicycle–car collisions: an in-depth study. *Accident Analysis & Prevention*, 30(5), 657-666.
- Schepers, J., Kroeze, P., Sweers, W., & Wüst, J. (2011). Road factors and bicycle–motor vehicle crashes at unsignalized priority intersections. *Accident Analysis & Prevention*, 43(3), 853-861.
- Schoner, J., & Levinson, D. (2013). Which Station? Access Trips and Bike Share Route Choice.
- Schuessler, N., & Axhausen, K. W. (2009). Map-matching of GPS traces on high-resolution navigation networks using the Multiple Hypothesis Technique (MHT).
- Segadilha, A. B. P., & da Penha Sanches, S. (2014). Identification of Factors that Influence Cyclists' Route Choice. *Procedia-Social and Behavioral Sciences*, 160, 372-380.
- Sener, I. N., Eluru, N., & Bhat, C. R. (2009). An analysis of bicycle route choice preferences in Texas, US. *Transportation*, 36(5), 511-539.
- Shaheen, S., Guzman, S., & Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia: past, present, and future. *Transportation Research Record: Journal of the Transportation Research Board*(2143), 159-167.
- Stinson, M., & Bhat, C. (2003). Commuter bicyclist route choice: analysis using a stated preference survey. *Transportation Research Record: Journal of the Transportation Research Board*(1828), 107-115.
- Train, K. E. (2009). *Discrete choice methods with simulation*: Cambridge university press.
- Winters, M., Davidson, G., Kao, D., & Teschke, K. (2011). Motivators and deterrents of bicycling: comparing influences on decisions to ride. *Transportation*, 38(1), 153-168.
- Wood, J. M., Lacherez, P. F., Marszalek, R. P., & King, M. J. (2009). Drivers' and cyclists' experiences of sharing the road: Incidents, attitudes and perceptions of visibility. *Accident Analysis & Prevention*, 41(4), 772-776.

APPENDIX

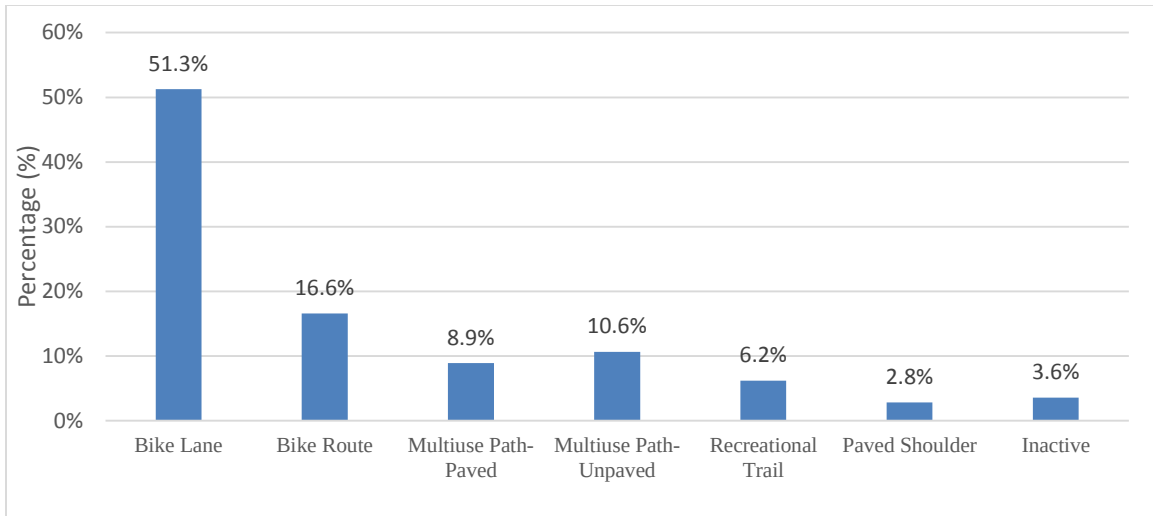


Figure 14 Categories of Bicycle-Specific Facilities in Phoenix, AZ

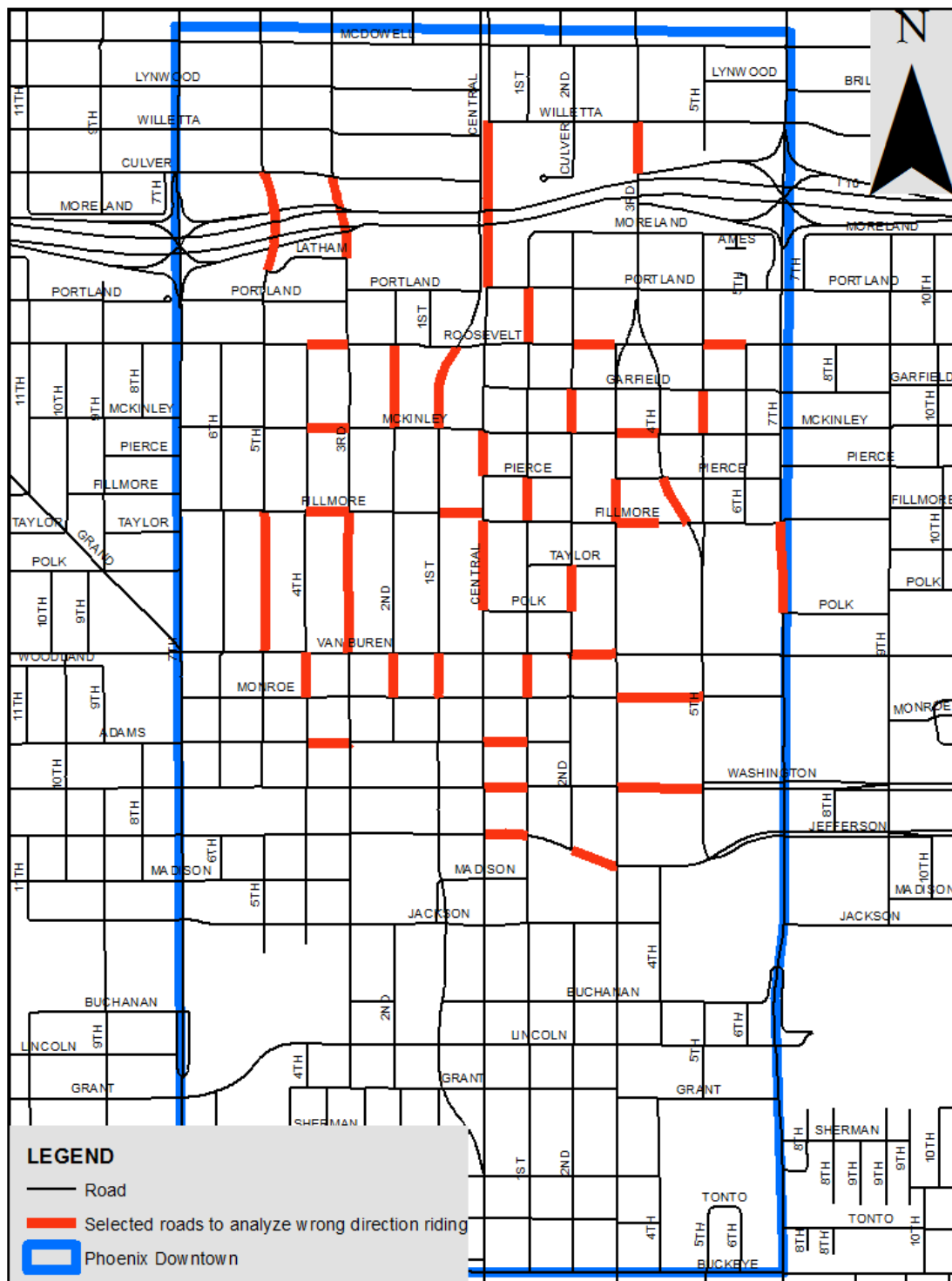


Figure 15 Street Segments Used for Analyzing Wrong Direction Riding Behavior

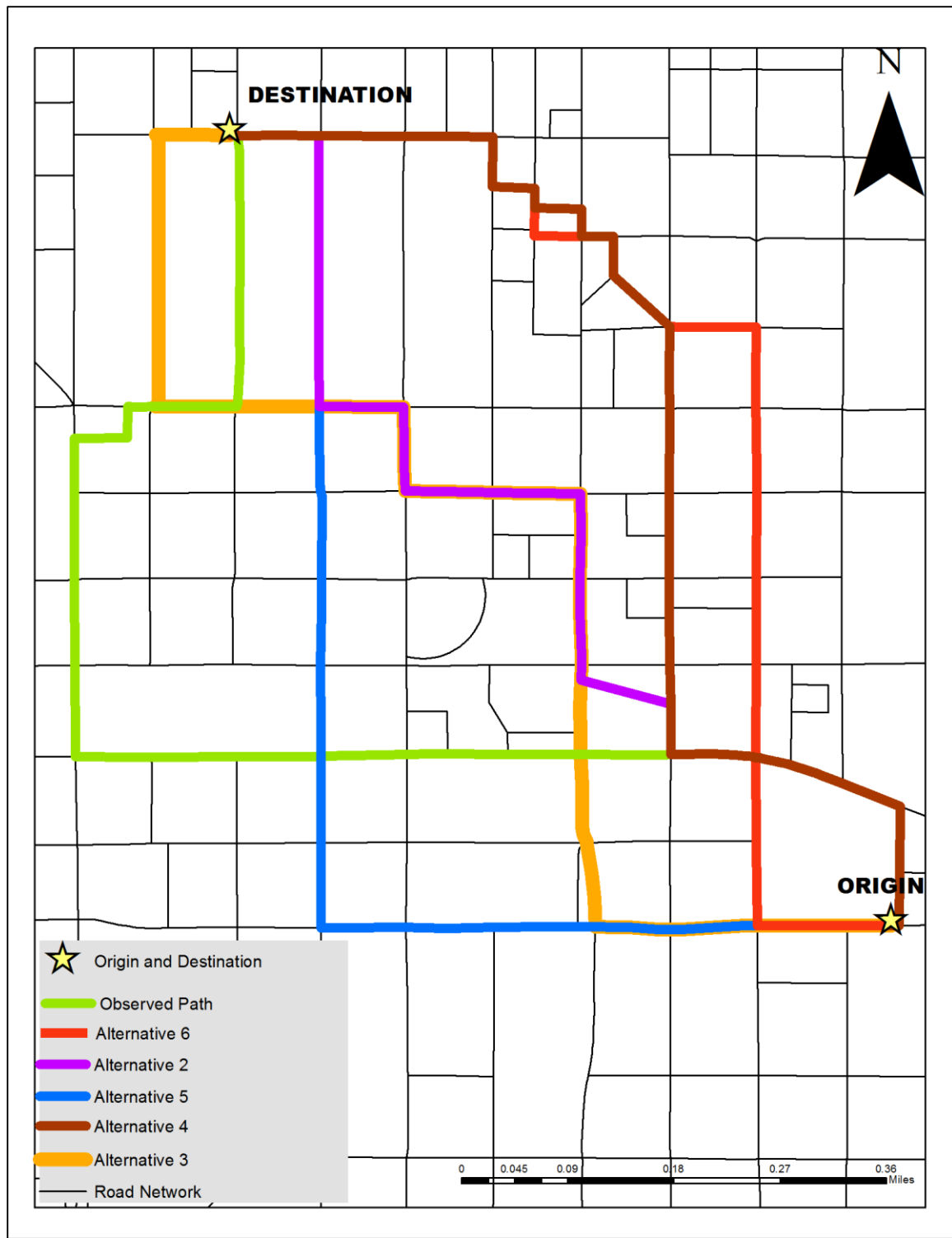


Figure 16 Set of Alternatives for a Pair of Origin and Destination

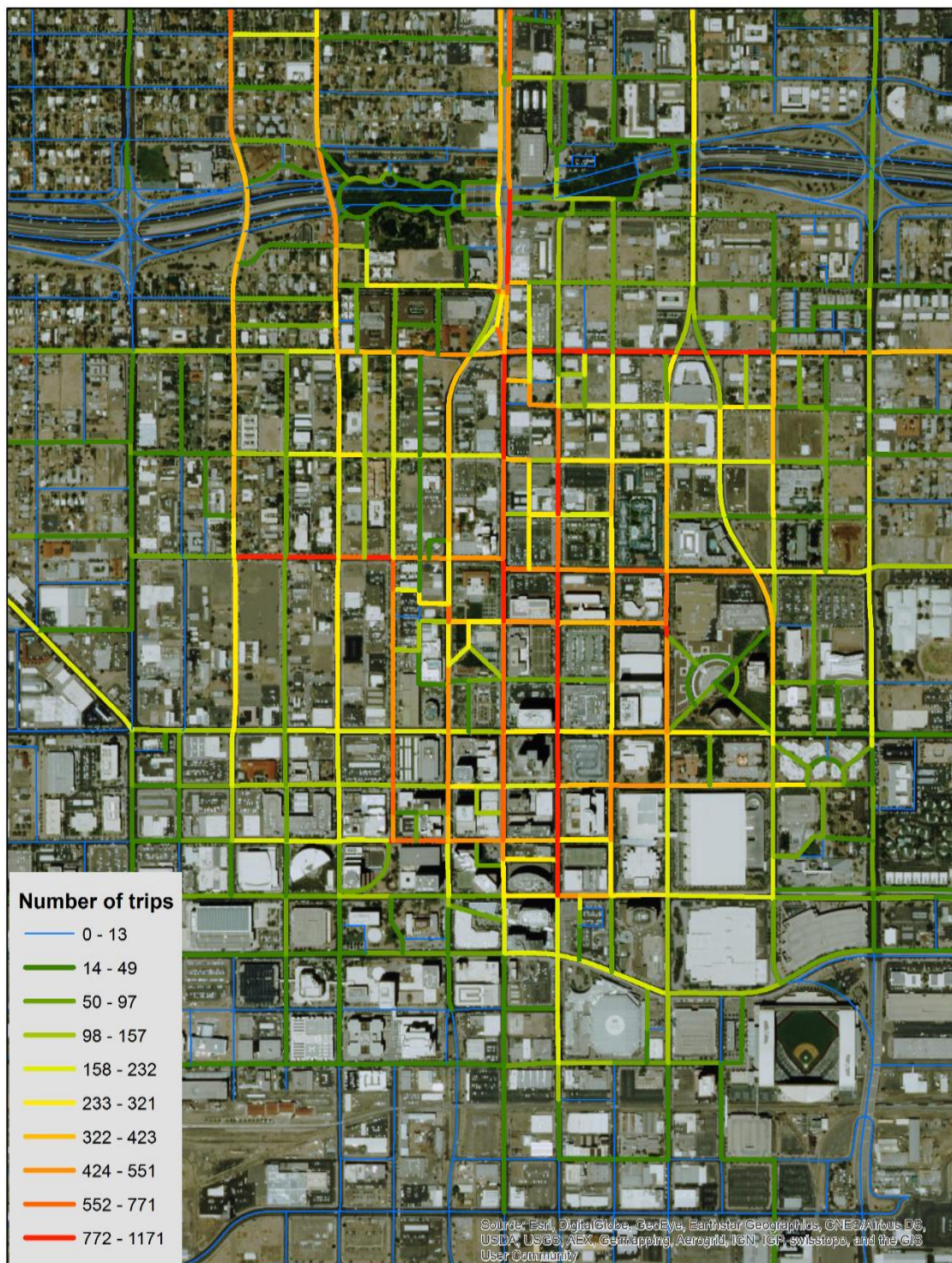


Figure 17 Volume of Bikeshare Trips on Streets of Downtown Phoenix

VITA

Ranjit Khatri received the B.S. degree in Civil Engineering from Pulchowk Engineering Campus, Lalitpur, Nepal in 2012. After his graduation, he worked as a Civil Engineer for Sitara Consult Pvt. Ltd. He moved to University of Tennessee, Knoxville for graduate studies in August 2014. After moving to the US for graduate studies, Ranjit started working under Dr. Christopher R. Cherry as Graduate Research Assistant on a project title “New technologies to assess bicycle safety”. His thesis is completely based upon this project. The project team also developed the smartphone app “I Bike KNX” for the city of Knoxville. He graduated with a Master of Science degree in Civil Engineering from the University of Tennessee in December 2015.